

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2001, 09:08:45 ; Search time 22.23 Seconds
(without alignments)
1219.025 Million cell updates/sec

Title: US-09-596-958-2

Perfect score: 2310

Sequence: 1 MSITLNNNTSSSPGLFQSG.....LGDVENHYKVPMSANLKVAE 447

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

/Database :

A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2310	100.0	447 20 AAW96260	Hypersensitive res
2	2310	100.0	447 21 AAY71094	Erwinia amylovora
3	2310	100.0	447 21 AAY84855	A hypersensitive r
4	559	24.2	424 20 AAW97851	Hypersensitive res
5	559	24.2	424 21 AAY71098	Pseudomonas syring
6	559	24.2	424 21 AAY84859	A hypersensitive r
7	359.5	15.6	197 19 AAW83014	Bacillus sp strain
8	359.5	15.6	197 19 AAW77412	Bacillus sp. pecti
9	341	14.8	221 20 AAY28446	Bacillus lichenifo
10	181.5	7.9	2411 21 AAB23860	Haemophilus influe
11	179	7.7	730 21 AAG29582	Arabidopsis thalia

12	179	7.7	752	21	AAG29581	Arabidopsis thalia
13	174	7.5	2042	19	AAW56319	Haemophilus paraga
14	165.5	7.2	344	18	AAW06600	Hypersensitive res
15	165.5	7.2	344	19	AAW75865	Pseudomonas solana
16	165.5	7.2	344	19	AAW62457	Pseudomonas solana
17	165.5	7.2	344	19	AAW61116	Hypersensitive res
18	165.5	7.2	344	20	AAW87641	A hypersensitive r
19	165.5	7.2	344	21	AAW71099	Pseudomonas solana
20	165.5	7.2	344	21	AAW84860	A hypersensitive r
21	164.5	7.1	898	18	AAW31853	Mycobacterium tube
22	157.5	6.8	2870	21	AAW95559	Caenorhabditis ele
23	157.5	6.8	3178	21	AAW95556	Caenorhabditis ele
24	156	6.8	2039	19	AAW56322	Haemophilus paraga
25	155.5	6.7	2353	17	AAW93933	Haemophilus adhesi
26	155	6.7	385	15	AAW45751	Erwinia amylovora
27	155	6.7	495	19	AAW39747	Human marCOSR prot
28	151	6.5	385	18	AAW05598	Hypersensitive res
29	151	6.5	403	19	AAW75865	Erwinia amylovora
30	151	6.5	403	19	AAW62455	Erwinia amylovora
31	151	6.5	403	19	AAW61114	Hypersensitive res
32	151	6.5	403	20	AAW87639	A hypersensitive r
33	151	6.5	403	21	AAW71093	Erwinia amylovora
34	151	6.5	403	21	AAW4854	A hypersensitive r
35	151	6.5	889	19	AAW68208	M. catarhalis str
36	150	6.5	520	19	AAW39748	Human marCOSR prot
37	150	6.5	520	21	AAW44332	Human PRO1559 prot
38	150	6.5	871	21	AAW95550	Chlamydia pneumoni
39	148.5	6.4	885	21	AAW90238	Mature Chlamydia a
40	148.5	6.4	928	21	AAW90237	Chlamydia antigen
41	148	6.4	440	21	AAW99354	Human PRO1411 (UNQ
42	148	6.4	440	22	AAW87551	Human PRO1411. Ho
43	148	6.4	440	22	AAW31205	Amino acid sequenc
44	148	6.4	440	22	AAW66103	Protein of the Inv
45	147.5	6.4	403	21	AAW58351	Lung cancer associ

ALIGNMENTS

RESULT 1
AAW96260
ID AAW96260 standard; Protein; 447 AA.
AC
XX AAW96260;
XX
DT 14-JUN-1999 (first entry)
XX
DE Hypersensitive response eliciting protein (HrpW).
XX

KW Hypersensitive response elicitor protein; hypersensitive response;
KW hrpW; pathogen; infection; crop protection; disease resistance;
KW pest resistance; transgenic plant; colouration; maturation.
XX
XX Erwinia amylovora.

OS Erwinia amylovora.

XX WO9907208-A1.

XX 18-FEB-1999.

PF 27-JUL-1998; 98WO-US15547.

XX 06-AUG-1997; 97US-0055108.

PR (CORR) CORNELL RES FOUND INC.

XX Beer SV, Kim JF;

XX WPI; 1999-167126/14.

DR N-PSDB; AAX09007.

XX New Erwinia amylovora hypersensitive response eliciting gene and
PT protein - useful for providing transgenic plants and seeds with
PT enhanced growth, and insect and disease resistance

XX PS Claim 1; Page 50-51; 54pp; English.

XX CC The hypersensitive response eliciting protein (hrp) or polypeptide

XX CC is produced as part of an active defense by plants against

XX CC incompatible pathogen infections. The hypersensitive response is a

XX CC rapid localised necrosis. The hrp protein and gene when used in

XX CC nucleotide constructs are useful for providing disease resistance to

XX CC plants, insect control to plants, and enhancing plant growth

XX CC (enhancing fruit size and earlier colouration and maturation), by

XX CC direct application of the protein to plants, or by producing

XX CC transgenic plants or seeds using the hrp gene.

XX SQ Sequence 447 AA;

Query Match 100.0%; Score 2310; DB 20; Length 447;

Best Local Similarity 100.0%; Pred. No. 3.1e-152;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSILTLNNNTSSPGLFQSGDNGLGHNANSALGCOQPIDRTQTEQMAQLLAELLSKLS 60

Db 1 msiltlnntsspglfqsgdnglgghnansalgqpidrtqteqmaqllellkslls 60

Qy 61 POSGNAATGAGNDQTTGCVGNAGLNGRKTAGTTPQSDSQNMLSEMNGLDQAITPDG 120

Db 61 pqsgnaatgagndqttgvgnagglngkrktagttppqsdsgnmlsemngngldqaitpdg 120

Qy 121 OGGGQIGDNPLLKAMKLIARMDGQSDQFGQPGTGNNSASSTSSGGSPFNDLSGGKA 180

Db 121 ogggqigdnplllkamklliarmdgqsdqfgpgtgnnsasstssggspfnldlsggka 180

Qy 181 PSGNSPQSNYSFVSTFSPSTPTSPDLFPSPPTKAAGGSTPTVTHDPVPVGSAGIGAG 240

Db 181 psgnspqsnysfvsfssptstpspldfpsptkaaggstptvdhdpvgsagigag 240

Qy 241 NSVAFTSAGANQTVLHDTITVKAGOVFGKGTFTAGSELGDDGGOSEKQPLFIEDGAS 300

Db 241 nsavftsaganqtvlhdtitvkagovfdgkgtftagselegdggsgenqkplfiedgas 300

Qy 301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360

Db 301 lknvtmgddgadgihlygdakidnlhvtnvgedaitvknpsagkkshveitnssfehasd 360

Qy 361 KILQLNADTNLSVDNVKAKDFGTFTVTRNGGGQGNWDLNLSHISAEDGKFSFVKSDSEGLN 420

Db 361 kilqlnadtnlsvndvknakdfgtftvtrngggqgnwdlnlshisaedgkfsfvksdsegl 420

Qy 421 VNTSDISLGDVENHYKVPMSANLKVAE 447

Db 421 vntsdislgdvenhykvpmsanlkvae 447

RESULT 2

AA71094

ID AA71094 standard; Protein; 447 AA.

XX AC AA71094;

XX DT 08-SEP-2000 (first entry)

XX DE Erwinia amylovora hypersensitive response elicitor #2.

XX KW Hypersensitive response elicitor; environmental stress resistance;

XX KW plant.

XX OS Erwinia amylovora.

XX PN WO200028055-A2.

XX PD 18-MAY-2000.

XX PF 04-NOV-1999; 99WO-US26039.

XX PR 05-NOV-1998; 98US-0107243.

XX PA (EDEN-) EDEN BIOSCIENCE CORP.

XX PI Wei Z, Schading RL;

XX DR WPI; 2000-376566/32.

XX DR N-PSDB; AAD00669.

XX PT Application of a hypersensitive response elicitor protein to plants to

XX PT impart stress resistance -

XX PS Disclosure; Page 10-12; 84pp; English.

XX CC The patent discloses a method to impart stress resistance to plants by

XX CC applying a hypersensitive response elicitor in a non-infectious form to

XX CC a plant or seed. The present sequence is a hypersensitive

XX CC response elicitor protein from Erwinia amylovora. The protein is

XX CC heat stable, protease sensitive and suppressed by inhibitors of plant

XX CC metabolism. It is used to impart stress resistance to plants.

XX SQ Sequence 447 AA;

Query Match 100.0%; Score 2310; DB 21; Length 447;

Best Local Similarity 100.0%; Pred. No. 3.1e-152;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSILTLNNNTSSPGLFQSGDNGLGHNANSALGCOQPIDRTQTEQMAQLLAELLSKLS 60

Db 1 msiltlnntsspglfqsgdnglgghnansalgqpidrtqteqmaqllellkslls 60

Qy 61 POSGNAATGAGNDQTTGCVGNAGLNGRKTAGTTPQSDSQNMLSEMNGLDQAITPDG 120

Db 61 pqsgnaatgagndqttgvgnagglngkrktagttppqsdsgnmlsemngngldqaitpdg 120

Qy 121 OGGGQIGDNPLLKAMKLIARMDGQSDQFGQPGTGNNSASSTSSGGSPFNDLSGGKA 180

Db 121 ogggqigdnplllkamklliarmdgqsdqfgpgtgnnsasstssggspfnldlsggka 180

Qy 181 PSGNSPQSNYSFVSTFSPSTPTSPDLFPSPPTKAAGGSTPTVTHDPVPVGSAGIGAG 240

Db 181 psgnspqsnysfvsfssptstpspldfpsptkaaggstptvdhdpvgsagigag 240

Qy 241 NSVAFTSAGANQTVLHDTITVKAGOVFGKGTFTAGSELGDDGGOSEKQPLFIEDGAS 300

Db 241 nsavftsaganqtvlhdtitvkagovfdgkgtftagselegdggsgenqkplfiedgas 300

Qy 301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360

Db 301 lknvtmgddgadgihlygdakidnlhvtnvgedaitvknpsagkkshveitnssfehasd 360

Qy 361 KILQLNADTNLSVDNVKAKDFGTFTVTRNGGGQGNWDLNLSHISAEDGKFSFVKSDSEGLN 420

Db 361 kilqlnadtnlsvndvknakdfgtftvtrngggqgnwdlnlshisaedgkfsfvksdsegl 420

Qy 421 VNTSDISLGDVENHYKVPMSANLKVAE 447

Db 421 vntsdislgdvenhykvpmsanlkvae 447

RESULT 3

AA784855

ID AA784855 standard; Protein; 447 AA.

XX AC AA784855;

XX DT 08-AUG-2000 (first entry)

XX DE A hypersensitive response elicitor protein.

XX KW Hypersensitive response; insect control; disease resistance;

hypersensitive response elicitor; plant growth; vegetable; crop;
ornamental plant.

Erwinia amylovora.

WO200020452-A2.

13-APR-2000.

05-OCT-1999; 99WO-US23181.

05-OCT-1998; 98US-0103050.

(BDEN-) EDEN BIOSCIENCE CORP.

Wei Z, Fan H, Niggemeyer JL;

WPI: 2000-303745/26.

N-PSDB; AAA14939.

Hypersensitive response elicitor polypeptides useful for imparting enhanced growth, disease resistance and insect resistance to plants, especially vegetables and ornamental flowers -

Disclosure; Page 12-13; 100pp; English.

The present sequence represents a hypersensitive response elicitor polypeptide. The specification describes hypersensitive response elicitor polypeptide fragments, which do not elicit a hypersensitive response. Instead, the proteins impart disease resistance to plants, enhance plant growth, and/or control insects. The polypeptide fragments may be used to these properties to plants. The plants which may be treated in this way include vegetables, crops and ornamental plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower, peanut, corn, potato, sweet potato, bean, pea, chlicory, lettuce, endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower, broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon, citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco, tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia, petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.

Sequence 447 AA;

Query Match 100.0%; Score 2310; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 3.1e-152; Mismatches 0; Indels 0; Gaps 0;
Matches 447; Conservative 0;

QY 1 MSILTLNNNTSSPGFLFQSGDNGLAGHNANSALGQOPIDRQTEQMAQLLAELKSLLS 60
|||||
DB 1 msiltilnntsspgflfqsqgnglaghnansalgqpdrqtieqmaqlaelklls 60
|||||

QY 61 PQSGNAATGAGNDQTTGVGNAGLNGRKCTAGTTPQSDSQNNLSENGNGLDQATPDG 120
|||||
DB 61 pqsgnaatgagndqttgvgnaglngnrkctagttpqdsdsqnnlsemngngldqatp 120
|||||

QY 121 QGGQIGDNPILLKMLKLIARMDGSDQFGQPCGTGNNSASSGTSSSGGSPFNDLSCGA 180
|||||
DB 121 qggqigdnplllkmlklliarmdgsgdqfgqpctgnnsassgtsssggspfn 180
|||||

QY 181 PSQNSPGCNTSPVSTSPPTSPPLDFPSSPTKAAGSTPVTDHPDVPVGSAGIGAG 240
|||||
DB 181 psgnspgcntspvstspptsppldfpssptkaagstpvtdhpdvpvgsagigag 240
|||||

QY 241 NSVAFTSAGANQTVLHDTITVAKGVDFGQGTFTAGSELGDCGQSNOKPLFILEDGA 300
|||||
DB 241 nsvaftsaganqtvvlhdtitvakgvdfgqgtftagselgdcgqsnokplfiled 300
|||||

QY 301 LKNVTMGDDGADGTHLYGDAKIDNLHVTNVGEDAITVKNPSACKKSHVEITNSSFEH 360
|||||
DB 301 lknvtmgddgadgthlygdakidnlhvtnvgedaitvknpsackkshveitnssfe 360
|||||

QY 361 KIQLNADTNLSVDNKKAKDFGTFRVTNGSQGNWDLNLSHSAEDGKFSFKVSDSEGLN 420
|||||

DB 361 kiqlnadtnlsvdnkvkakdfgtfrvtngsqgnwdlnlsishaedgkfsfvksdsegl 420
|||||

QY 421 VNTSDISLGDVENHYKVPMSANKVAE 447
|||||

DB 421 vntsdislgdvenhykvmsankvae 447
|||||

RESULT 4
AAW97851
ID AAW97851 standard; Protein: 424 AA.
XX
AC AAW97851;
XX
DT 07-JUN-1999 (first entry)
XX
DE Hypersensitive response elicitor HrpW.
XX
KW Hypersensitive response elicitor; dspE gene; HrpW;
XX transgenic plant; disease resistance; insect resistance.
XX
OS Pseudomonas syringae pv. tomato.
XX
FH Key Location/Qualifiers
FT Domain 1..186
FT /note= "hypersensitive response elicitor-like
FT domain"
FT Region 119..186
FT /note= "region of 6 imperfect glycine-rich repeats"
FT Domain 187..424
FT /note= "C-terminal domain"
XX
PN WO9907207-A1.
XX
PD 18-FEB-1999.
XX
PF 24-JUL-1998; 98WO-US15501.
XX
PR 06-AUG-1997; 97US-0055107.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Alfano JR, Charkowski A, Collmer A;
XX
DR WPI: 1999-167125/14.
DR N-PSDB; AAX24347.
XX
PT New hypersensitive response eliciting (dspE) gene and protein
PT useful for providing transgenic plants and seeds with enhanced
PT growth, and insect and disease resistance
XX
PS Claim 18; Page 38-39; 56pp; English.
XX
CC This the amino acid sequence of hypersensitive response element
CC (HRE) HrpW of Pseudomonas syringae pv. tomato DC3000, as deduced
CC from the nucleotide sequence of an isolated DNA molecule (see
CC AAX24347). HrpW includes an HRE-like domain that is rich in Gln,
CC Ser and Gly and which includes 6 imperfect glycine-rich repeats
CC with many acidic and polar residues that align with similar
CC repeats in the HrpZ proteins of P. syringae pv. syringae and P.
CC syringae pv. tomato. The amino acid sequence of this region
CC suggests alternating beta-sheets and turns that may form a barrel
CC structure. The C-terminal domain of hrpW is similar to several
CC fungal and bacterial pel proteins. The HrpW protein or isolated
CC DNA molecule can be used to impart disease resistance to plants, to
CC enhance plant growth and/or to control insects on plants. This is
CC achieved by applying the HRE protein in a non-infectious form to
CC plants or plant seeds. Alternatively, transgenic plants or plant
CC seeds transformed with DNA encoding the HRE can be provided.
XX
SQ Sequence 424 AA;

XX PS Disclosure; Page 26-28; 100pp; English.

XX CC The present sequence represents a hypersensitive response elicitor

CC CC polypeptide. The specification describes hypersensitive response

CC CC elicitor polypeptide fragments, which do not elicit a hypersensitive

CC CC response. Instead, the proteins impart disease resistance to plants,

CC CC enhance plant growth, and/or control insects. The polypeptide

CC CC fragments may be used to these properties to plants. The plants which

CC CC may be treated in this way include vegetables, crops and ornamental

CC CC plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,

CC CC peanut, corn, potato, sweet potato, bean, pea, chichory, lettuce,

CC CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,

CC CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,

CC CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,

CC CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,

CC CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,

CC CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.

XX SQ Sequence 424 AA;

Query Match 24.2%; Score 559; DB 21; Length 424;
 Best Local Similarity 36.1%; Pred. No. 4.8e-31;
 Matches 147; Conservative 54; Mismatches 132; Indels 74; Gaps 13;

QY 37 QPTDROTIEQMAOLLAEELKSL---LSPQSGNAATGAGGNDQTTGVCNAGGLNGRKGTAG 93

Db 72 kpdssd--niaklisalmsllqmlcnsnkqdtncqpdspqfngnglg----- 122

QY 94 TTPQSDQNMLSEMGNGLDQAITPDQGGGQIGDNPPLKAMLKLIARMDGSDQDFGQP 153

Db 123 -tpsads-----gggg-----tpdatgggg-gdtp-----satggg 151

QY 154 GTGNNSASSGTSSTGGGFFNDLGGKAPSGNSGVSTPSPPTSPSPPLDFPS 213

Db 152 ggdtpatggggggggtptatggg---sggtptatgggggvtqitpql-----a 200

QY 214 SPTKAAGGSTPVTDPHPDVGSGAGTGNVAFVTSAGANQVLTHTITVKAQGVFDGKQT 273

Db 201 nprtsg-----tgsvsdtag-----teaqklnvkvdtlkvgagevfdghgat 245

QY 274 FTAGSELGQGQSQNKPLFILEDASLKNVTMGDDGADGHLHYG-----DAKIDNLHVTN 329

Db 246 ftdaksmngdggngqkpmfelaegatlkvnlgenevvgilhvkaaknaqevtldnvhqan 305

QY 330 VGEDAITVKNPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNG 389

Db 306 vgedlitvkggegaavcnlnknssakgddkvvqlnathklkldnfkaddfgumvrtng 365

QY 390 GQQ-GNNDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLGDVENHY 435

Db 366 gkqfdmsielngleanhgkalfalvksdsddklkatgntamtdvkhay 412

RESULT 7

AAW83014

ID AAW83014 standard; Protein; 197 AA.

XX AC AAW83014;

XX DT 28-JAN-1999 (first entry)

XX DE Bacillus sp strain KSM-P15 pectic acid lyase.

XX KW Bacillus sp. strain KSM-P15; pectic acid lyase; protopectinase;

XX KW protopectin; polygalacturonic acid; detergent; cotton; surfactant;

XX KW cellulase; protease; bleaching agent.

XX OS Bacillus sp.

XX PN W09845393-A2.

XX

PD 15-OCT-1998.

XX PF 08-APR-1998; 98WO-JP01613.

XX PR 08-SEP-1997; 97JP-0242736.

XX PR 09-APR-1997; 97JP-0091142.

XX PA (KAOS) KAO CORP.

XX PI Hatada Y, Ito S, Kasai M, Kobayashi T, Koike K;

PI Shikata S, Suzumatsu A, Tsumadori M, Wada Y;

XX DR N-PSDB; AAV69879.

XX DR WPI; 1998-568339/48.

XX PT Detergent composition containing protopectinase active at alkaline

PT pH - on protopectin and polygalacturonic acid, provides better

PT removal of muddy soil

XX PS Example; Page 71-72; 80pp; English.

XX CC The present invention describes a detergent composition which contains

CC a protopectinase having an optimum pH 7 or higher against protopectin

CC and polygalacturonic acid substrates. Inclusion of protopectinase

CC gives a composition that provides better removal of muddy soil,

CC particularly from socks. The present sequence represents pectic acid

CC lyase from Bacillus sp. strain KSM-P15, which is used in an example

CC from the present invention. Pectic acid lyase exhibits protopectinase

CC activity.

XX SQ Sequence 197 AA;

Query Match 15.6%; Score 359.5; DB 19; Length 197;
 Best Local Similarity 46.3%; Pred. No. 1.2e-17;
 Matches 81; Conservative 29; Mismatches 56; Indels 9; Gaps 4;

QY 253 TVLHDTITVRKAGQVDFGKGTFTAG-SELGQGQSENGKPLFILEDASLKNVTMGDDGA 311

Db 3 tvvhetirvpagqtdgkgtqyvanpntlgdgsaqenqkplfleagaslknnvvgapaa 62

QY 312 DGHLHYDADKIDNLHVTNMGEDAITVKNPNSAGKKSHVEITNSSFEHASDKILQLNADTNL 371

Db 63 dgvhcyctctitnviwedgedaitlk--ssgt---vnlsggaaaykaykvfginaagti 117

QY 372 SVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFVKSDSE---GLNVNT 423

Db 118 nlrnfdraddlgklvrqnggttykvmmhvenchrsrvkdalrtdsstgrlvnt 172

RESULT 8

AAW77412

ID AAW77412 standard; Protein; 197 AA.

XX AC AAW77412;

XX DT 07-JAN-1999 (first entry)

XX DE Bacillus sp. pectic acid lyase.

XX KW Bacillus sp. KSM-P15; pectic acid lyase; pectinic acid lyase; pectin;

XX KW detergent; food-processing; fibre-processing agent; cell wall.

XX OS Bacillus sp.

XX PN EP870834-A1.

XX PD 14-OCT-1998.

XX PF 09-APR-1998; 98EP-0106586.

XX PR 08-SEP-1997; 97JP-0242735.

XX PR 09-APR-1997; 97JP-0091142.

```
XX PA (KAOS ) KAO CORP.
XX PI Hatada Y, Ito S, Kobayashi T, Koike K, Suzumatsu A;
XX PI Yoshimatsu T;
XX PR WPI: 1998-523159/45.
DR N-PSDB; AAV59478.
XX PT New Bacillus pectic acid lyase - useful as a detergent component, a
XX PT food-processing agent and a fibre-processing agent
XX PS Claim 1; Page 16-17; 29pp; English.
XX CC The present sequence represents a pectic acid lyase isolated from
XX CC microorganism Bacillus sp. KSM-P15. The pectic acid lyase has high
XX CC pectic acid lyase activity which degrades pectin in plant cell walls
XX CC and fibre in vegetables, and so is useful as a component of detergents,
XX CC a food-processing agent, or a fibre-processing agent. The pectic acid
XX CC lyase has a higher optimum reaction pH (10.3-10.7) than known Bacillus
XX CC pectic acid lyases (pH 8-9.5) and so has wider industrial applications.
XX CC Unlike present pectic acid lyases, the new enzyme has a high enzyme
XX CC activity, and can be produced on a mass scale.
XX SQ Sequence 197 AA;
Query Match 15.6%; Score 359.5; DB 19; Length 197;
Best Local Similarity 46.3%; Pred. No. 1.2e-17;
Matches 81; Conservative 29; Mismatches 56; Indels 9; Gaps 4;
QY 253 TVLHDTITVKGQVFGKQGTFTAG-SELGSGQSENQKPLFILEDGASLKNVMTGDDGA 311
Db. 3 tvvhtirvpagqtdfgkgtvyvnpntltdgsgaenqkplfrleagaslknvvgapaa 62
QY 312 DGHLYGDAKIDNLHVTNVGDAITVKPNSACKKSHVEITNSFEHASKILQLNADTNL 371
Db 63 dgvhcygddctlnvldwedgedaltik--srgt---vnlsaggaaaykaydkvfqinaagtl 117
QY 372 SDVNKAKDGFCTFVRTNGGQGNWDLNLSHISAEDGKFSEVKSQDSP--GLNVNT 423
Db 118 nlrnfraddlgkrlvqnggttykvvmvncnsvrkdailrtdsstdgrivnt 172
RESULT 9
ID AAY28446
XX AAY28446 standard; Protein; 221 AA.
XX AC AAY28446;
XX DT 17-JAN-2000 (first entry)
XX DE Bacillus licheniformis Pectate lyase I.
XX KW Pectate lyase I; EC 4.2.2.2; pectin degrading enzyme; ATCC 14580;
KW catalytic active domain; cellulose binding domain; CBD; operably linked;
KW optimum activity; pH; detergent composition; yarn; cellulosic fibre;
KW recycled waste paper; pulp; retting process; animal feed; wine; juice;
KW transgenic plant.
XX OS Bacillus licheniformis.
XX FT Key Location/Qualifiers
FT Peptide 1..27
FT Protein /label= Pro-sequence
FT 28..221
FT /label= Mature_pectate_lyase_I
FT Misc-difference 133
FT /note= "Conserved residue"
FT Misc-difference 155
FT /note= "Conserved residue"
XX PN W09927083-A1.

XX 03-JUN-1999.
XX PD 24-NOV-1998; 98WO-DK00514.
XX PF 24-NOV-1997; 97DK-0001344.
XX PR 06-MAY-1998; 98US-0073684.
XX XX (NOVO ) NOVO-NORDISK AS.
XX PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;
XX DR WPI; 1999-610578/52.
XX DR N-PSDB; AAX89484.
XX XX New isolated pectate lyase enzymes -
XX PT Claim 6; Page 79-80; 93pp; English.
XX PS The present sequence is pectate lyase I, which is a pectin degrading
XX CC enzyme derived from Bacillus licheniformis, ATCC 14580. The enzyme
XX CC comprises a catalytic active domain and a cellulose binding domain
XX CC (CBD) that are operably linked to each other. It shows optimum activity
XX CC at pH greater than 9 and temperature 55 degree centigrade. It can be used
XX CC in detergent compositions, for cleaning hard surfaces, for machine
XX CC treatment of fabrics, for improving the properties of cellulosic fibres,
XX CC yarn, woven or non-woven fabric, for the degradation of plant material
XX CC e.g. recycled waste paper, mechanical paper-making pulps or fibres
XX CC subjected to retting process, for preparing animal feed and for
XX CC processing wine or juice. DNA encoding the enzyme can also be used for
XX CC the production of transgenic plants.
XX SQ Sequence 221 AA;
Query Match 14.8%; Score 341; DB 20; Length 221;
Best Local Similarity 41.7%; Pred. No. 2.6e-16;
Matches 73; Conservative 33; Mismatches 59; Indels 10; Gaps 2;
QY 254 VLHDTITVKGQVFGKQGTFTAGSELGSGQSENQKPLFILEDGASLKNVMTGDDGADG 313
Db 31 vvhtkivvekgqtdgkgrliagpelgdsqredqkpfkvedgatlknnvvlgapaadg 90
QY 314 IHLYGDAKIDNLHVTNVGDAITVKPNSACKKSHVEITNSFEHASKILQLNADTNLSV 373
Db 91 vhtygnasinnvwwedvgedalrvk----segsvtlnggsarlaadkifginkastftv 145
QY 374 DNVKAKDGFCTFVRTNGGQGNWDLNLSHISAEDGKFSEVKSQDSPSEGLNVNTSDISL 428
Db 146 knftadggkflrqjggstfkavvndhctlnmkealftrds-----stssvtm 195
RESULT 10
ID AAB23860
XX AAB23860 standard; Protein; 2411 AA.
XX AC AAB23860;
XX DT 17-JAN-2001 (first entry)
XX DE Haemophilus influenzae adhesin (Hia) protein from type c strain API.
XX KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX OS Haemophilus influenzae.
XX PN W0200055191-A2.
XX PD 21-SEP-2000.
```


PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.7%; Score 179; DB 21; Length 730;
Best Local Similarity 23.2%; Pred. No. 0.00019;
Matches 115; Conservative 52; Mismatches 158; Indels 170; Gaps 24;


```
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145270.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160788.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.78; Score 179; DB 21; Length 752;
Best Local Similarity 23.28; Pred. No. 0.0002;
Matches 115; Conservative 52; Mismatches 158; Indels 170; Gaps 24;

Qy 9 NTSSSPGLFQSGDNGLGCHNANSALGOOPIDROTIEQMAQLLAELLKSLLSPOSSNAAT 68
Db 154 stgakpgasgigsdsgsig-----sagtnpbadgtre-----teknaggskpssgsagt 202
Qy 69 GAG-----GNDQTTGVGNAGLNGRKGTTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGG 124
Db 203 npgasavngete--knaggskpssgsagtnpga-----sagnggeteknvggskpssg 254
Qy 125 QIGDNPLLKAMKLIARMWDQSDQFGQPGT-----GNNSASSGT----- 164
Db 255 kagtnp-----ganagnggteknaggskssgsgeartnpbasagnggetv 299
Qy 165 -----SSSGSPFNLSGGKAPSGNSGNSPVSTFSPSTPTSPSTPLDFSSPTK 217
Db 300 snigdtessnaggsksdngannagsgiesnag--stgtnfgaggtggigdtessdaggsaktn 357
Qy 218 AAGGSTPVTDPHPVGSAGIGAGNSVAFTSAGANQTVLHDITVKAGQVFDGKGGOTFTAG 277
Db 358 sgnggt--nd-----gasgigsdngstgtnpgag-----ggttdsnleg 393
Qy 278 SELGDGGOSEKQKPLFILEDGAS-LKNVTMGDDGADGTHLYGDAKIDNLHVTNVGEDAIT 336
Db 394 tennvggketn-----pgasgign-----sdgstgtspeg-----tesnadg 430
Qy 337 VKPNSAGKKSHVEITNSSFEHASDKILOLNADTNLSVDNVKAKDFGTFTVRTNGGQQGN-W 395
Db 431 tknttgkkesn---tgs-----esntsnspqkleaq-----gngngnqw 466
Qy 396 DLNLSH-----ISAEKGFSPVKSD--SEG-----LNVTNTSISL 428
Db 467 ddgtdhdgvmkjhavavglgieqirfdyvkngqlkegpfhgvkgrggtstieishpdeyl 526
Qy 429 GDVENHYKVPMSANL 443
Db 527 vsveglyd---ssnl 538

RESULT 13
AAW56319
ID AAW56319 standard; Protein; 2042 AA.
XX
AC AAW56319;
XX
DT 19-AUG-1998 (first entry)
XX
DE Haemophilus paragallinarum antigenic protein #1.
XX
```

QEDATVRFNSAQ KMSHVEITINSSEHASSDRILEQENADINLSVD 37

[illegible]

Result No.	Score	Query %		Length	DB	ID	Description
		Match	Length				
1	2310	100.0	447	2	T18447		HrPW protein - Erw
2	353.5	15.3	221	2	A70045		pectate lyase homo
3	233.5	10.1	219	2	S68364		pectate lyase (EC
4	207	9.0	1660	2	A70869		hypothetical glyci
5	199.5	8.6	1306	2	A70934		hypothetical glyci
6	193	8.4	694	2	F70868		hypothetical glyci
7	188	8.1	639	2	D70931		hypothetical glyci
8	186.5	8.1	591	2	B70523		hypothetical glyci
9	186	8.1	1329	2	E70917		hypothetical glyci
10	185.5	8.0	1381	2	E70806		hypothetical glyci
11	185.5	8.0	1489	2	D70807		hypothetical glyci
12	183.5	7.9	242	2	A45724		pectate lyase (EC
13	183	7.9	588	2	F70971		hypothetical glyci
14	182	7.9	606	2	F70816		hypothetical glyci
15	182	7.9	731	2	C70974		hypothetical glyci
16	181.5	7.9	484	2	G70846		hypothetical glyci
17	180.5	7.8	1079	2	B70807		hypothetical glyci
18	179	7.7	730	2	F96559		hypothetical prote
19	179	7.7	1538	2	F70846		hypothetical glyci
20	179	7.7	3016	2	S77300		hypothetical prote
21	178	7.7	741	2	G70917		hypothetical glyci
22	177.5	7.7	1901	2	F70806		hypothetical glyci
23	177.5	7.7	2232	2	T34434		hypothetical prote
24	176.5	7.6	562	2	B70953		hypothetical glyci
25	175	7.6	645	2	F70825		probable PPE prote
26	174	7.5	615	2	H70589		hypothetical glyci
27	172.5	7.5	584	2	G70804		hypothetical glyci
28	172	7.4	3507	2	T34513		hypothetical prote
29	171.5	7.4	667	2	A70893		hypothetical glyci

```
|||||
Db 421 VNTSDISLGDVENHYKVPMSANLKVAE 447

RESULT 2
A70045
pectate lyase homolog yvpA - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: A70045
R:Kunst, F.; Ogasevara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.: Bron, S.; Brouillet, S.; Pruthi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.: Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Togononi, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: A70045
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-221 <KUN>
A:Cross-references: GB:299121; GB:AL009126; NID:92635827; PIDN:CAB15500.1; PID:ell186183;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvpA

Query Match 15.3%; Score 353.5; DB 2; Length 221;
Best Local Similarity 41.8%; Pred. No. 5.6e-13;
Matches 76; Conservative 31; Mismatches 70; Indels 5; Gaps 1;

QY 250 ANOTVLHDTITVKAGVDFKGTFTAGSELGDSQENQKPLFILEDGASLKNVTMGDD 309
Db 27 ADKVVHETIIIPKNTTYDGKQRFVAGELGDSQENQDPFRVEDGATLKNVVLGAP 86

QY 310 GADGHIYGDADINLHVNTNGEDAITVKPNSAGKKSHVETITNSSFEHASDKILQINADT 369
Db 87 AADGVHTYGNVNTQNKWEDVGEDALTVK----REGKVTIDGSAQAQKADKIFQINKAS 141

QY 370 NLSVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSVKSDESEGLNVTSDISLG 429
Db 142 TETVKNFTADNGGKFIQLGSGTFHVDVIIDKCTITNMKEAIFRTDSKTSTVRMTNTRY 201

QY 430 DV 431
Db 202 NV 203

RESULT 3
S68364
pectate lyase (EC 4.2.2.2) C - fungus (Fusarium solani)
C:Species: Fusarium solani
C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
C:Accession: S68364
R:Guo, W.; Gonzalez-Candelas, L.; Kolattukudy, P.E.
Arch. Biochem. Biophys. 323, 352-360, 1995
A:Title: Cloning of a new pectate lyase gene pelc from Fusarium solani f. sp. pisi (Nect
A:Reference number: S68364; MUID:96063610
A:Accession: S68364
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-219 <GUO>
A:Cross-references: EMBL:U13049
C:Genetics:
```

```
A:Gene: pelc
A:Introns: 52/1; 102/3
C:Keywords: carbon-oxygen lyase

Query Match 10.1%; Score 233.5; DB 2; Length 219;
Best Local Similarity 30.1%; Pred. No. 2.3e-06;
Matches 63; Conservative 39; Mismatches 72; Indels 35; Gaps 6;

QY 220 GGSTPVTGHPDPVGSAGICAGNSVAFTSAGANQTVLHDTITVKAGOVDFGKQGTAGSG 279
Db 8 GGVPRKPTDH-----ISNS-----XVIEVKAGQVYDGRKWKYDRSGG 43

QY 280 LGDG-QGSENQKPLFILEDGASLKNVTMGDDGADGHIYGLYGDADINLHVNTNGEDAITVK 338
Db 44 ACKQNEGDKDAVELLHEGATLKNVIGKQDSGVCHCKGHCTLEFVWFEDVCEDAISIK 103

QY 339 PMSACKSHVETITNSSFEHASDKILQINADTNLSVDNVKAKDFGTFVRT--NGGQGNWD 396
Db 104 EDKAGKESW--IIGGGAYHASDKVYQHNGCGVTNINFEVDYGLYRSCGNCQCKRN 161

QY 397 LNLSHISAEDGKFSVKSDESEGLNVTSD 425
Db 162 VYIEGVTAKE-----GELAGINANYGD 184

RESULT 4
A70869
hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70869
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genc
A:Reference number: A70500; MUID:98295987
A:Accession: A70869
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1660 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:93261507; PIDN:CAAL16067.1; PID:g27?
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2490c
C:Superfamily: collagen alpha I(IV) chain

Query Match 9.0%; Score 207; DB 2; Length 1660;
Best Local Similarity 28.1%; Pred. No. 0.00077;
Matches 91; Conservative 19; Mismatches 130; Indels 84; Gaps 15;

QY 7 NNNTSSPLFQSGDNGLGGHNANSALGQQPFDRTQTEQMAQLLAELLKLLSP--OSG 64
Db 476 NGGAGGNGGLV---GDGGAGGHGGDGAAG-----AGVADMTAIFLGSSTPGEDGG 523

QY 65 N----AATCAGG-NDQTTGVGNAGLNGRKCTAGTTPQSDSONMLSENNCLDOA--IT 117
Db 524 NGGAGGAGGAGAHAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 583

QY 118 PDGCGGGTGNPLLLKMLKLIARMMDQSDQDFGQPGTGNNSASSGTSSSGG---SPND 174
Db 584 GDGGAGGAGGADAPAGRA-----GSGQVGGDGGAGGAGGAGGAGGAGGAGGAGGAGG 633

QY 175 LSGGKAPSGNSP-SCNYSFVSTFSPSPPTPTSPPLDPFSPSTKAAGGSTPVTDHPDVG 233
Db 634 GDGGAGGDDGDPGAGGKGGAG-----GAGATEGVTGATGATVHSG--G 674

QY 234 SAGIGAGNSVAFTSAGANQTVLHDTITVKAGOVDFGKGTFTAGSELGDSQENQKPLF 293
Db 675 NGGKG-GNGADATVAGAN-----GKGAGGAGNGGLVGGGAG----- 710
```

QY 294 ILEDG----ASLKNVTMGDDGADG 313
||| | :|:|:|:|
Db 711 --GDGSGAAGANVGEDGADG 732

RESULT 5
A70934
hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70934
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987
A: Accession: A70934
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1306 <COL>
A: Cross-references: GB:AL021942; GB:AL123456; NID: g3242298; PIDN: CAAL17449.1; PID: g290963
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: Rv0578c
C: Superfamily: collagen alpha 1(IV) chain

Query Match 8.6%; Score 199.5; DB 2: Length 1306;
Best Local Similarity 26.0%; Pred. No. 0.0015;
Matches 88; Conservative 16; Mismatches 122; Indels 113; Gaps 11;

QY 8 NWTSSPGLFQSGDNGLGHNANALGQOPIDRQIEQAQLLELLKSLSPQSGNAA 67
||| | :|:|:|:|
Db 754 NGVAGSQGPGGAGGTGCGVGGNGRGIDGADGAT----- 789

QY 68 TGAGGNDQTTGVGNAGLGKRGKTAGTTTQSDSQNNLSEMGNNGLDQAITPDGQGGGQIG 127
||| | :|:|:|:|
Db 790 --AGARGQDGGAGGAGKGGGTGCP-----GGAGPAGTGTGSGGAGNG 832

QY 128 DNPLLKAMKLKLIARMMDGSDQFGPGCTGNNSASCTSSGSGFPFNDLSGGKA-PSGNP 186
||| | :|:|:|:|
Db 833 -----GSGGTGGDPGDGNGANGSVFTNNGIGGNGGNGAGNAPSGAGG 875

QY 187 SGNYSPVSTFPSTPTSPSL-----DFPSSPTKAG-----C 221
||| | :|:|:|:|
Db 876 SGAG--STFG-----ATGSSSIHVNGGNGGNGDHALSGNAAAGGNGGNSLRG 929

QY 222 STPVTDPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGTFTAGSELG 281
||| | :|:|:|:|
Db 930 SGGAGCHGGNGGNASRGMGDGTGCGAGN-----AQICNG-----GAGGNG 973

QY 282 DGQSQNQKPLFL-----EDGSLKNVTMGDDGADG 313
||| | :|:|:|:|
Db 974 DGTGSDGNPGAITGSGRGGDGGVGGGSGVAGDGDG 1012

RESULT 6
F70868
hypothetical glycine-rich protein Rv2487c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70868
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987
A: Accession: F70868

A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-694 <COL>
A: Cross-references: GB:AL021246; GB:AL123456; NID: g3261507; PIDN: CAAL16064.1; PID: g279
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: Rv2487c
C: Superfamily: elastin

Query Match 8.4%; Score 193; DB 2: Length 694;
Best Local Similarity 25.4%; Pred. No. 0.0016;
Matches 86; Conservative 28; Mismatches 140; Indels 84; Gaps 14;

QY 18 QSGDNGLGCH-NANALGQOPIDRQIEQAQLLELLKSLSPQSGNATCAGNDQT 76
||| | :|:|:|:|
Db 256 QTGCDGTGCHGTAGTPTGCTGDGAT-----ATAGSKRATGAGGDCGT 300

QY 77 TGVGNAGLNGRKGTAGTTTQSDSQNNLSEMGNNGLDQAITPDGQGGQIGDNPLLKAML 136
||| | :|:|:|:|
Db 301 AAAGGGGNGDGGVA---QGDIAFAFGDGGSDGVAAGSGGGGAGGG---AFV 352

QY 137 KLARMDQSDQFGPGCTGNNSASCT-SSSGSPFNDLSGCK--APSGNSPSGNYSPV 193
||| | :|:|:|:|
Db 353 HIATATSTGSGGFG--GNGAASAGAGGAGGAGGLLFGDGGNGAGGAGGI 410

QY 194 STFPSTPTSPSLDFPSSPTKAAAGGSTPVT--DHPDP-----VGSAG 236
||| | :|:|:|:|
Db 411 GG-----DGATGGPGSGGNAGIARFSPDPEAPDVVGGKGGGKGGSG 456

QY 237 IGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGTFTAGSELGDSQENKPLFILE 296
||| | :|:|:|:|
Db 457 LGVGGAGGTGGAGNG-----GAGGLFCNGGN---GGNAGAGG-----D 493

QY 297 DGASLKNVTMGDDGADGCIHLIGDAKIDNL-HVTNVED 333
||| | :|:|:|:|
Db 494 GGAGVAGVGGNGGGGTATFHEDPVAGVWAVGGVGGD 531

RESULT 7
D70931
hypothetical glycine-rich protein Rv1803c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: D70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987
A: Accession: D70931
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-639 <COL>
A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAAL17724.1; PID: e125
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: Rv1803c
C: Superfamily: unassigned collagens

Query Match 8.1%; Score 188; DB 2: Length 639;
Best Local Similarity 22.9%; Pred. No. 0.0027;
Matches 96; Conservative 33; Mismatches 144; Indels 146; Gaps 17;

QY 7 NNNTSSPGLFQSGDNGLGHNANALGQOPIDRQIEQAQLLELLKSLSPQSGNA 66
||| | :|:|:|:|
Db 159 NGCGGSGAGLWGGGCGGCGAGANGAQP-----GK 191

QY 67 ATGAGGNDQT-----TGVGNAGLNGRKGTAGTTTQSDSQNNLSEMGNNGLDQAITPDGQ 122
||| | :|:|:|:|

Db 192 AGSGGNGAGGWIYHGHHGGAGGNGNA-TAPGASAGFDGAGGNG-----GSGGRG 245
Qy 123 GQIGDNPLLKAMLLIARMMDQSDQFQPGCTGNNSASSGTSSSGGSPFND----- 174
Db 246 GLLFGNG-----GNSVGMGCGQTNDTAGDSAGSGGLGGNGCAQGGW 290
Qy 175 LSGGKAPSGNSPSGNYSPVSTFSPSTPTSPDLDFSSPTKAAAGSTPTVTDHPDPVGS 234
Db 291 LIGNGGGQDGSAGG-----GTDSTGTGVNM-----GASGGAGTAGNGGDAGL 334
Qy 235 AGIG-----AGNSVAFTSAGANQTVLHDTITVKAGQVFDGKQFTTAGSELGDGQSQENQ 289
Db 335 VNGGAGGNGNGAAGSALGTT-----IFGSGGVGGSGDGGNGG----- 375
Qy 290 KPLFTLEDGASLKNVTMGDDGADGHIHYGDAKIDNLHVTNVGEDAITVKPNKSAKKSHVE 349
Db 376 ---WLFSGSAS-----CGNGCGQ-----GDA-----GTNGFAGFGSAGSGGGWVG 412
Qy 350 ITNSSFEHASKILOLNADTNLSVDNVKAKDFGTFRVTRNGGQGN-WDLNLHSHISAEDG 407
Db 413 AVN-----FGPISVQGFGLP--GHGGDGGNGDVGAGSLSIQFG 449

RESULT 8
B70523
hypothetical glycine-rich protein Rv0297 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000
C:Accession: B70523
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: B70523
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-591 <COL>
A:Cross-references: GB:296800; GB:AL123456; NID:g3261800; PIDN:CAB09596.1; PID:e321655;
A:Experimental source: strain H37RV
C:Genetics:
C:Superfamily: unassigned collagens

Query Match 8.18; Score 186.5; DB 2; Length 591;
Best Local Similarity 23.18; Pred. No. 0.0029;
Matches 91; Conservative 31; Mismatches 119; Indels 153; Gaps 15;

Qy 6 LNNNTSSPGLFQSGGDNGL-----GHHNANSALGQPIDROTIEQMAQLLAELKSLSP 61
Db 116 IGNANGAPGTGQAGDGLLFGNGGCGAPGQ-----AG 152

Qy 62 QSGNAA-----TCAGNDQTTGVGNAG----- 83
Db 153 GAGGAAGFFGNGGNGDGGAGAGGAGGTAGWFFGFGNGAGGIGVAGINGLGGAGGD 212

Qy 84 -----GLNKRGTAG-----TTQSDSNMLSEMNGLDQAITPDGOGG 123
Db 213 GGNAGFFGNGGNGGAGAGVNAVNPGLATPTVPAAN-----GGNGLNLVGPVGTAGG 267

Qy 124 GQIGDNPLLKAMLLIARMMDQSDQFQPGCTGNNSASSG-----TSSSGGSPFNDLSG 177
Db 268 GADGAN-----GSAIGQAG--GAGDGGNASTSGGIGTAQTGGAGA--GGAGG 312

Qy 178 KAPSGNSPSGNSPVSTFSPSTPTSPDLDFSSPTKAAAGSTPTVTDHPDPVGSAGI 237
Db 313 DCAPGGNGG-----GSVEHTGATGSSASGNGATGNGGVGAPG- 353

Qy 238 GAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKQFTTAGSELGDGQSQENQPLFLED 297

Db 354 GAGNGGHHVSGGSVNT-----AGA--GKGNGCTGGAGGPGGH----- 390
Qy 298 GASLKNVTMGDDGADGHIHYGDAKIDNLHVTNVG 331
Db 391 GGSVLSPGVDSGNGGAGDGGAGVSATDIAGTG 424

RESULT 9
E70917
hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70917
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: E70917
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1329 <COL>
A:Cross-references: GB:295844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:g21310
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1450c
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 8.18; Score 186; DB 2; Length 1329;
Best Local Similarity 25.28; Pred. No. 0.0084;
Matches 86; Conservative 25; Mismatches 156; Indels 104; Gaps 14;

Qy 7 NNTSSPGLFQSGDNGLGGHNANSALGQGP-----IDROTIEQMAQLLAELKLS 57
Db 704 NGANATTFG--AKGDDGGHGGPAGAGNGGGGPGGLAGNLFQNGIGQVGGSGGKGAG 761

Qy 58 LLSFQSGNAAT-----GAGNDQTTGVGNAGGLN---GRKGTAGTTPOSQ 101
Db 762 GLAGDGGANGNFAFGDNGHGGNGNPGAGGGGAGSTPGAKGAGFTPTSGD 821

Qy 102 NMLSEMNGLDQAITPDGQGGQIGDNPPLLKAMLLIARMMDQSDQFQPGGTGNNSSAS 161
Db 822 G--GDGNGNSQVVGNGGDDGGNGG-----GSAGTGGNGR 858

Qy 162 SGTSSSGSPENDLSGK-APSGNSPSGNSPVSTFSPSTPTSPDLDFSSPTKAG 220
Db 859 GGDGAFGMSANATNPGENPGN-PCGN-----CGAG 890

Qy 221 GSTPTVDHPDPVGSAGI-----GAGNSVAFTSAGANQTVLHDTITVKA-GOV--FDGKGQ 272
Db 891 CA-----GGAGLNGGNGAGGNGGLGFGGNGAAGANGVAVGAPGPGGAGGHHG 940

Qy 273 TTTAGSELGDGQSQENQKPLFLEDGASLKNVTMGDDGADG 313
Db 941 AGGNGGAGGCGQ-----VSDGAGGAGGAGDGGCAPG 974

RESULT 10
E70806
hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70806
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987

A:Accession: E70806
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1381 <COL>
A:Cross-references: GB:AL0202022; GB:AL123456; NID:g3261554; PID:CAAI7744.1; PID:g292444
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3507
C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.0%; Score 185.5; DB 2; Length 1381;
Best Local Similarity 25.0%; Pred. No. 0.0094;
Matches 79; Conservative 25; Mismatches 141; Indels 71; Gaps 11;

QY 19 SGGDNGLGHNANSALGQQ--PIDRTIQMAQLLAELLKSLSPSGNAATGAGNDQT 76
DB 779 SGGDGGKGGGGGGTGGSCAPI-----GGAGGTGGSGGH 814
QY 77 TGVGNAGLNGRKGTAGTTTPQSDQNMLSEMNGNLQDQAITPDGQGG-GQIGDNPLLKAM 135
DB 815 AGKGGAGGI-GAQTITITVPGNGN--AGDGGNGNAGA---GGNGSGDFGNT----- 863
QY 136 LKLIARMMDGQSDQFGQPTGNNNSASSTSSCGSPFNDLSGKAPS-GNSPSGNTSPVS 194
DB 864 -----TSGASGSGGNGGAGTGGTGLSGGNGGNGGNGGNGGNGG 913
QY 195 TFSPPSTPTSPPLDFPSSPTKAAAGSTPVTDHPDVPVGSAGIGAGNSVAFTSAGANQTV 254
DB 914 AHCTVGAQFVPATSLTPNGAGGNGGTSNGGAPGAPGPTTGGNAGSQIGDGGN 973
QY 255 LHDTITVKAQVFDG-----RGQTFTAGS-----ELGDDGQSQNQKPLFILED 297
DB 974 GGD-----GKGGGDGADAVNVVEMPTPQAATGTAGSAGDPTGGNGGPGTTPGPMVAPP 1028
QY 298 GASLKNVTMGDDGADG 313
DB 1029 PTPITVQVQGGGGAG 1044

RESULT 11
D70807
hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1489 <COL>
A:Cross-references: GB:AL0202022; GB:AL123456; NID:g3261554; PID:CAAI7751.1; PID:g292445
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3514
C:Superfamily: collagen alpha 1(IV) chain

Query Match 8.0%; Score 185.5; DB 2; Length 1489;
Best Local Similarity 23.9%; Pred. No. 0.01;
Matches 85; Conservative 26; Mismatches 136; Indels 109; Gaps 12;

QY 10 TSSSPGLFOSGDNGLGGHNANSALGQPIDRTIQMAQLLAELLKSLSPSGNAATG 69
DB 726 TTGNAGVGGGSGAGGTNGSGAGG-----TDGCGGAGGAG 763
QY 70 AGGNDQTTGVG-----NAGGLNGRKGTAGTTTPQSDQNMLSEMNGNLQDQAIT 117

DB 764 GAGADNPITGIGTGGDGTGAAGAGGAGGAAGTGTG-----GGMIGTTGNAGV----- 812
QY 118 PDQGGGGQTDGNPLLKAMKLIARMMDGQSDQFGQPG-TGNNSASSTSSSGSPFNDLS 176
DB 813 --GGAGGGGGGG-----AGGAGADADQPCATGCTGTGAGGAGGAGGGSSCA 858
QY 177 GKKAPSGNSPSGNYSPVSTFSPSTPTSPDLPSSPTKAAAGSTPVTDHPDVPVGSAG 236
DB 859 GGTNGSGGA--GGTCGQV-----VAGGAGISFSGNSNGTGGTGGVGGTGGDGNAG 908
QY 237 IGA-----GNSVAFTSAGAN-----QTVLHDTITVYKAG 264
DB 909 TGAGDPKGGTGTGTGGSGGAGGSGGANFNGGTGTGTGCKGLNTDGLSATSGTG 968
QY 265 QV--FDGKGGTFTAGSEL-----GDGQSQNQKPLFILEDGAELKNVTMGDDGADG 313
DB 969 GTGGTGGKGGTGGAGDDSAGGTGGTGAGGAGGAGGLANTGGTAGNAGIGGGGGG 1024

RESULT 12
A45724
pectate lyase (EC 4.2.2.2) - fungus (Fusarium solani)
C:Species: Fusarium solani
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A45724
R:Gonzalez-Candelas, L.; Kolattukudy, P.E.
J: Bacteriol. 174, 6343-6349, 1992
A:Title: Isolation and analysis of a novel inducible pectate lyase gene from the phyt
A:Reference number: A45724; MUID:93015682
A:Accession: A45724
A:Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-242 <CON>
A:Cross-references: GB:M94691; NID:gl68155; PIDN:AAA33338.1; PID:gl68156
A:Experimental source: isolate T8
A:Note: sequence extracted from NCBI backbone (NCBIN:115473, NCBI:115474)
C:Keywords: carbon-oxygen lyase

Query Match 7.9%; Score 183.5; DB 2; Length 242;
Best Local Similarity 30.5%; Pred. No. 0.0014;
Matches 62; Conservative 32; Mismatches 80; Indels 29; Gaps 8;

QY 235 AGTGAGNSVAFT-----SAGANOTVLHDTITVKAQGVFDCKGOTFTAGSELGDDGQSQENQ 289
DB 9 AALVGTSSAAVTTLPKSACA--TSPTAVPVKGS--YDGMKRFEPKVPKVGQDETGE 64
QY 290 K-PLFILEDGAELKNVTMGDDGADGTHLYGDAKIDNLHVTNVGDEATITVKNPSAGKSHV 348
DB 65 KDAFFILENGATLSNVIIGASQAEVHCKGTCTLLNNVMWADVCEDAVTLKOTGTS---- 120
QY 349 EITNSFEHASDKILQNLADTNLSVONVAKDFGTFTVFT-----NGGQGGNLDNLSHI 402
DB 121 YINGGGAFFASDKIIIFNGRGTVHVVDYAEADYKGLSRSCGCKNDNGPR---NVIVENS 177
QY 403 SAEDGKRFSEVKSDSEGLNVNTSD 425
DB 178 VAVDGGVLC-----GINTNYGD 194

RESULT 13
F70971
hypothetical glycine-rich protein Rv3367 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Feb-2000
C:Accession: F70971
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

Search completed: October 18, 2001, 09:09:33
Job time: 48 sec

842-1-2410

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2001, 09:08:45 ; Search time 12.52 Seconds
(without alignments)
735.135 Million cell updates/sec

Title: US-09-596-958-2
Perfect score: 2310
Sequence: 1 MSILTLNNTSSSPGLFQSG.....LGDVENHYKVPMSANLKVAE 447

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	559	24.2	424	US-09-120-817-2	Sequence 2, Appli
2	359.5	15.6	197	US-09-402-668-2	Sequence 2, Appli
3	341	14.8	221	US-09-198-936-4	Sequence 4, Appli
4	190.5	8.2	62	US-09-402-668-10	Sequence 10, Appli
5	165.5	7.2	344	US-08-891-254-7	Sequence 7, Appli
6	165.5	7.2	344	US-08-819-539-7	Sequence 7, Appli
7	165.5	7.2	344	US-09-030-270A-7	Sequence 7, Appli
8	165.5	7.2	344	US-08-984-207-7	Sequence 7, Appli
9	165.5	7.2	344	PCT-US96-08819-7	Sequence 7, Appli
10	160.5	6.9	907	US-09-010-928B-4	Sequence 4, Appli
11	156.5	6.8	1912	US-08-409-995-4	Sequence 4, Appli
12	156.5	6.8	1912	US-08-685-467-4	Sequence 4, Appli
13	155.5	6.7	2353	US-09-377-155-33	Sequence 33, Appli
14	155.5	6.7	2353	US-08-913-942-4	Sequence 4, Appli
15	155	6.7	385	PCT-US93-06243-2	Sequence 2, Appli
16	155	6.7	495	US-09-794-795-2	Sequence 2, Appli
17	155	6.7	495	US-09-249-200-2	Sequence 2, Appli
18	151	6.5	385	US-08-891-254-3	Sequence 3, Appli
19	151	6.5	385	US-08-819-539-3	Sequence 3, Appli
20	151	6.5	385	PCT-US96-08819-3	Sequence 3, Appli
21	151	6.5	403	US-08-200-724A-2	Sequence 2, Appli
22	151	6.5	403	US-09-030-270A-3	Sequence 3, Appli
23	151	6.5	403	US-08-851-376A-2	Sequence 2, Appli
24	151	6.5	403	US-08-984-207-3	Sequence 3, Appli
25	150.5	6.5	745	US-09-010-928B-28	Sequence 28, Appli
26	150.5	6.5	870	US-09-010-928B-2	Sequence 2, Appli
27	150	6.5	520	US-08-794-795-6	Sequence 6, Appli

28	150	6.5	520	4	US-09-249-200-6	Sequence 6, Appli
29	147.5	6.4	489	2	US-08-794-795-7	Sequence 7, Appli
30	147.5	6.4	489	4	US-09-249-200-7	Sequence 7, Appli
31	147.5	6.4	518	1	US-08-392-367B-2	Sequence 2, Appli
32	147.5	6.4	518	3	US-08-893-467A-2	Sequence 2, Appli
33	145	6.3	334	4	US-09-060-756-728	Sequence 728, App
34	144	6.2	1581	4	US-09-110-517-2	Sequence 2, Appli
35	143.5	6.2	679	4	US-08-913-942-15	Sequence 15, Appli
36	138	6.0	353	2	US-08-687-702-37	Sequence 37, Appli
37	137	5.9	300	3	US-08-765-856-2	Sequence 2, Appli
38	137	5.9	300	4	US-08-935-009A-2	Sequence 2, Appli
39	137	5.9	842	5	PCT-US96-02331-15	Sequence 15, Appli
40	136.5	5.9	1477	1	US-08-038-682-4	Sequence 4, Appli
41	136.5	5.9	1477	1	US-08-302-832-4	Sequence 4, Appli
42	136.5	5.9	1477	2	US-08-530-198-4	Sequence 4, Appli
43	136.5	5.9	1477	2	US-08-469-880-4	Sequence 4, Appli
44	136.5	5.9	1477	2	US-08-728-470-4	Sequence 4, Appli
45	136.5	5.9	1477	2	US-08-617-697-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-120-817-2
; Sequence 2, Application US/09120817
; Patent No. 6172184
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Charkowski, Amy
; APPLICANT: Alfano, James R.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: PSEUDOMONAS SYRINGAE AND ITS USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Nixon, Hargrave, Devans & Doyle LLP
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,817
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,107
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1741
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-120-817-2

Query Match 24.2%; Score 559; DB 4; Length 424;
Best Local Similarity 36.1%; Pred. No. 1.1e-36;
Matches 147; Conservative 54; Mismatches 132; Indels 74; Gaps 13;

Db 194 DGGSGAGGAGGANGADGGNGVNGNQANGPQAGDVNGANGAD-----DGSEDDGSLTGVL 248

Qy 130 PLLKAMLKLIARMMD-----GSDQFGQPGTGNNSSAGSTSSGGSP--FNDLSGGK 179
| : : || | | | | | : | : | :
Db 249 QKLWKILNALVQMOMQGGGLGGNQAQGGSKAGNAPASGANPGCANOPGSAIDDOSSGO 306

```

RESULT 7
US-09-030-270A-7
; Sequence 7, Application US/09030270A
; Patent No. 5977060
; GENERAL INFORMATION:
; APPLICANT: Zitter, Thomas A.
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: INSECT CONTROL WITH A
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,270A
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,226
; FILING DATE: 28-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1521
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-030-270A-7

```

```

Query Match      7.2%; Score 165.5; DB 2; Length 344;
Best Local Similarity 32.0%; Pred. No. 1.3e-05;
Matches 57; Conservative 20; Mismatches 80; Indels 21; Gaps 7;

QY 18 QSGD---NGLGHNANSALGOQPIDQTTEMAQLLAEL-LKSLSPSQSNAATGAGN 73
      |||: |||: | | | : : : : |||: | | | : | | | |
Db 134 QPGNGKNGVGVGGANGAKGAGGOGGLAEALQEIQLAQLGGGGAGGAGGVGGAGGA 193
      |||: |||: | | | : : : : |||: | | | : | | | |
QY 74 DQTTGVGNAGGLRKGKTAGT--TPQSDSNMLSEMNGNLIDQAITPDG--QGSGQIGDN 129
      |||: |||: | | | : : : : |||: | | | : | | | |
Db 194 DGSAGGAGGAGGADGCGVNGVNGAQPQAGDVNGANGAD-----DGSDEQGLTGVL 248
      |||: |||: | | | : : : : |||: | | | : | | | |
QY 130 PLKAMUKLIARMD-----GSDQFGPGTGNNASSTGSSGGSP--FNDLSGKK 179
      |||: |||: | | | : : : : |||: | | | : | | | |
Db 249 QKLMLKILNALVQMQQGGGLGGGNAQGGSKGAGNAPGANGPCANPGASADDSGQ 306
      |||: |||: | | | : : : : |||: | | | : | | | |

```

RESULT 8
US-08-984-207-7
; Sequence 7, Application US/08984207
; Patent No. 6235974
; GENERAL INFORMATION:

APPLICANT: Qiu, Dewen
 APPLICANT: Wei, Zhong-Min
 APPLICANT: Beer, Steven V.
 TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
 TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 STREET: P. O. Box 1051, Clinton Square
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/984,207
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/033,230
 FILING DATE: 05-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldman, Michael L.
 REGISTRATION NUMBER: 30,727
 REFERENCE/DOCKET NUMBER: 19603/1201
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 344 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PS-08-984-207-7

```

Query Match      7.2%   Score 165.5; DB 4; Length 344;
Best Local Similarity 32.0%; Pred. No. 1.3e-05;
Matches 57; Conservative 20; Mismatches 80; Indels 21; Gaps 7;

QY    18 QSGD---NLGGHNANSALCOQPIDQTTEQAQLLAE-LKSLSPSGNAATCAGN 73
       ||| :||| | | | : : : :||| : | : ||||
Db    134 PQGNDKGNGVGGANGAKGGCGGLAEAIQETELAOLGGGAGAGAGGVGGAGGA 193
       ||| :||| | | | : : : :||| : | : ||||

QY    74 DOTTGVGNAGLNGRKGTAGT--TPQSDSQNMLSMGNGLDQAITPDG--GGGQIGDN 129
       ||| :||| | | | : : : :||| : | : ||||
Db    194 DGSAGGAGGAGANGADGGVNGNQANGPQAGDVNGACD-----DGSEDDGGILTVL 248

QY    130 PLLKAMKLIRAMD-----GSDQPGQPTGNNSASGTSSSGGSP--FNDLSGGK 179
       ||| :||| | | | : : : :||| : | : ||||
Db    249 QKLMLKINALVMQQMGQGLGGNOAQGGSKAGNASPASGANPNQGASDDQSSQG 306

```

RESULT 9
PCT-US96-08819-7
; Sequence 7, Application PC/TUS9608819
; GENERAL INFORMATION:
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.


```
;
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08819
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/475,775
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10051
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-08819-7

Query Match 7.2%; Score 165.5; DB 5; Length 344;
Best Local Similarity 32.0%; Pred. No. 1.3e-05;
Matches 57; Conservative 20; Mismatches 80; Indels 21; Gaps 7;

QY 18 QSGD---NGLGHNANSALGQPIDRTIEQMAQLLAEL-LKSLSPSGNNAATGAGN 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 QPGNDKNGVGGANGAKGAGGGGGLAEALQETIEQLAQLGGGAGGAGGAGGAGGA 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 74 DQPTGVNAGLNGRGTAGT--TPQSDSONLSEMNGLDQAITPDG--QGGGQIGDN 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 DGSAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 PLKAMKLIARMMD-----GQSDQFGQPGTGNNSASSGTSSSGSP--FNDLSGK 179
   . ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 QKLMKILNALVQMQQGLGGGQAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-010-928B-4

Query Match 6.9%; Score 160.5; DB 2; Length 907;
Best Local Similarity 21.4%; Pred. No. 0.00012;
Matches 85; Conservative 35; Mismatches 134; Indels 143; Gaps 14;

QY 12 SSPGLFQSGDNGLGGHNANSALGQPIDRTIEQMAQLLAELKSLSPSGNA--ATG 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 SGPGGYGGG-SCAGGTGPGGAG-----AGGAGGSGG 634
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 AGNDQTTGVNAGLNGRGTAGTTPQSD-----SQNMLSEMNGLDQ 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 AGSGGAGGSGGAGGSGGSGGTTTITDITDADGPITISELTISGAGSGPGG 694
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 115 A----ITPDGQGGQIGDNPLLKAMKLIARMMDGSDQFGPCTGNNSASSGTSSSG- 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 AGPGGVGPGSGPGGVGPG-----VSGPGG-VGPGGSGGSGGSGGPGGV 740
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 170 -----SPFNDLSGKAPSGNSPSGN-----YSPVSTFSP-----PST 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 741 GPGGYGPGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 800
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 PPSPTSLDFPSSPTKAGSGTPVTDHPDVSAGIAGNSVAFTSAGANQTVLHDTIV 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 801 PGSPGGAYG-PGSPGGAYPSSRPDVMVNGIMSAMOGSGFN----- 841
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 262 KAGQVFDGQFTAGSELGDSGQSENQKPLFLEDDGASLKNVTMGDDGDGDIHLYGDAK 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 842 ---QMFGNLSQYSSGS-----GTCNPNVNVLMAL 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 322 IDNLH-VTVNGDAITVKPNSACKKSHVEITNSFEH 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 871 LAALHCLSHGSSSFAPSPPTPAAMSAYSNSVGRMFAY 907
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Gene III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
```

REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

Query Match 6.8%; Score 156.5; DB 1; Length 1912;
Best Local Similarity 19.7%; Pred. NO. 0.00068;
Matches 103; Conservative 61; Mismatches 168; Indels 191; Gaps 22;

Qy	70	AGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQ-----NMUSEMGNGLDQAITPDG	120
Db	1209	SAGNKEITNVKSA--LKTYKDTQNTADETQDKEFHAAVKANAEYEFVKGNGATVSAKTDN	1266
Qy	121	QG-----GGOIGDNPLLKAMLLKLTARMMDGSDQFGQPGTGN-----SASSGTSSS	167
Db	1267	NGKHTVTIDVAEAKVGDG-----LEKOTDCKIKLKVDNTDGNLLTVDATKGASVA	1317
Qy	168	GGSPFNLDLGGKAPSGNSPGWSPVSTFSPSTPTSPSPLDPPSPSTKAAGSTPV-T	226
Db	1318	-----KGBFNAVTT---DATTAQCTNANERGKVVVKGSGNGATATET	1355
Qy	227	DHPD--PVGSGAGIGAGNSVAF-----TSAGANQTV-LHDTITVKAGOVF-	267
Db	1356	DKKKVATVGDVAKAINDAATFVKVVDSDSATIDDSPTDDGANDALKAXDILTLLKAGKNLK	1415
Qy	268	---DGKQTTTAGSE-----LGDGGOSENOKPLFILLEDGASL---KNVTMGD	308
Db	1416	VKRDGKNITFALANDLSVKRSATVSKLSLGTNGKNV-----ITSRTKGLNFAKDSKTGD	1470
Qy	309	DGADGTHLYG---DAKIDNLHVTNVGDAIT-----VKPN	340
Db	1471	DA---NIHLANGIASTLTDTLNLSGATNLGGNGITDNEKKXAASVKDVLNAGWNVGVKPA	1528
Qy	341	SAGKSHVEITNSSPEHASDKILQLNADTNLSVDNVKAKDFG-----	382
Db	1529	SANN---QVENIDEFATYDVTDFVSGBKDRTTSVTVESKDKRTEYKIGAKTSVIKDHN	1584
Qy	383	-----TFVRTNGGOGN-----WDLNLSHTSAEDBCKF	409
Db	1585	GKLFCTCKELKDANNQCVTTETDGDGDEBGLVTAKAVIDAVNKGARVKTGTGANGQNDFF	1644
Qy	410	SFVKS-----DSEGLNVNTSDISLGDVENHYKVPMSANLKV	445
Db	1645	ATVASCTNVTFADGNGTTAETVKANDGSTITVKNVKKVADGKLK	1687

```

RESULT 12
US-08-685-467-4
: Sequence 4, Application US/08685467
: Patent No. 6060059
: GENERAL INFORMATION:
: APPLICANT: St. Geme III, Joseph W.
: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PS-08-685-467-4

```

Query Match	6.8%	Score 156.5;	DB 3;	Length 1912;
Best Local Similarity	19.7%	Pred. No. 0.00068;		
Matches 103; Conservative	61;	Mismatches 168;	Indels 191;	Gaps 22;

Qy	70	AGNDQTGVGNAGGLNGRKGTAGTTPQSDSQ-----NWLSEMGNNGLDQAIPDG	120
Dq	1209	SAGNKEITNVKSA--LKTYKDQTADETQDRFEHAAVKNAVEFVGKINATVSAKTDN	1266
Qy	121	QG-----GGQIGDNPLLAMKLIAARMMDGSGDSFGPGTGNN---SASGSGTSSS	167
Dq	1267	NGRHTVTIDVAEAKVGDG-----LEKDDTGKILUKVDNTDGNLLTVDATKGASVA	1317
Qy	168	GGSPFNLDLGGKAPSGNSPSGKNVPSTPPTTSPLDFPSSPTKAAGSTPV-T	225
Dq	1318	-----KGFENAVTT---DATTAQGTANERGVVVKSGSNATATET	1355
Qy	227	DHPD---PVSAGISGAGNSVAF-----TSAGANTQV-LHDITTVKAGOVF-	267
Dq	1356	DKKKVATVGDVAKAINDAATFKVKNENDSATIDDPTDGDANDALKAXDTLTLRAGNLK	1415
Qy	268	---DGKGQTFTAGSE-----LGCGQSGENOKPLFILEDGASL---KNVTMGD	308
Dq	1416	VKRDKGNITFALANDLSVKSVATSVDKSLJGTNGKNVN---ITSDTKGLNFAKDSKTGD	1470
Qy	309	DGADGTHLYG-----DAKIDNLHVNTVCEDAIT-----VKPN	340
Dq	1471	DA---NIHLNGIASLTJDTLLNSGATTNLGGNGITDNEKKXAASVKDVLNAGNNRVGYPKA	1528
Qy	341	SACKSHVEITNSSPEHASDKLIQLNADTNLSVDNVNKAADF-----	382
Dq	1529	SANN---QVENIDEFATYDTPVDSGDKDTTSVTVESKDKGRKTEVRIKAGTSVIKHDN	1584
Qy	383	-----TFVRINGGOQH-----WDLNLSHISAEDCKE	409
Dq	1585	GKLFCTGKELKDANNNGVTVFETDGRDEGNGLVTAKAIVDYNKAGRKYKTTGCANGQNDPF	1644
Qy	410	SPVKS-----DSEGLANYNTSDISLGDOVENHYKVPMSANLKV	445
Dq	1645	ATVASGTNTVFDAGNCJTAEVTKANDGSTITVRYKNVADVAGLKL	1687

RESULT 13
US-09-377-155-33
; Sequence 33, Application US/09377155

Patent No. 6197312
GENERAL INFORMATION:
APPLICANT: PEAK, Ian Richard Anselm
APPLICANT: JENNINGS, Michael Paul
APPLICANT: MOXON, E. Richard
TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
FILE REFERENCE: 065064/0128
CURRENT APPLICATION NUMBER: US/09/377,155
CURRENT FILING DATE: 1999-08-19
PRIOR APPLICATION NUMBER: PCT/AU98/01031
PRIOR FILING DATE: 1998-12-14
PRIOR APPLICATION NUMBER: GB 9726398.2
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 33
LENGTH: 2353
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match 6.7%; Score 155.5; DB 4; Length 2353;
Best Local Similarity 19.7%; Pred. No. 0.0011;
Matches 103; Conservative 61; Mismatches 168; Indels 191; Gaps 22;

QY 70 AGGDDTTGVNAGGLNGRKGRTAGTTPOSDSQ-----NMLSEMNGNGLDQAITPDG 120
DB 1210 SAGNKEITNVKSA--LKYTKDONTADETQDKEFHAAVKNAVEFVGKNGATVSAKTDN 1267
QY 121 QG-----GGQIGDNPLLKAMKLIAARMMDGQSDQFGQPGTGN-----SASSGTSSS 167
DB 1268 NGKHTVTIDVAEAKVGDG-----LEKDTGKIKLVNDTGNLLTVDATKGASVA 1318
QY 168 GGSFFNDLGGKAPSGNSPVSTPSPPTSPSPSPPLDPPSPPKAAGSTPV-T 226
DB 1319 -----KGEFNAVTT---DATTAQGTNANERGVVYVKGSGNGATATET 1356
QY 227 DHPD--PVSAGIAGNSVAF-----TSAGANTV-LHDTITVKAGOVF- 267
DB 1357 DKKKVATGVDVAKAINDAATFVKVENDSDSATIDSDTDGANDALKAGDTLLKAGKNLK 1416
QY 268 ---DGKGQTFAGSE-----LGDGQSENQKPLFLEDCASL---KNVTMGD 308
DB 1417 VKRDGKNITFALANDLSVKVSATVSKLSLGTNGKNVN-----ITSOTKGLNFAKDSKTGD 1471
QY 309 DGADGTHLYG-----DAKIDNLHVTNVGEDAIT-----VKPN 340
DB 1472 DA--NIHLNGIASTLTDTLNLSGATTNLGGNGITDNEKKRAASVKOVNLNAGWNVGVKPA 1529
QY 341 SACKKSHVEITNSSFEHASKILQLNADTNLSVDNPKAKDFG-----WDLNLSHISAEDGKF 409
DB 1530 SANN---QVENIDFVATVDTVDVSGDKDTSVTVESKDNKGRTEVKIGAKTSVTKDHN 1585
QY 383 -----TFVRTNGGQGN-----WDLNLSHISAEDGKF 409
DB 1646 ATVASGNTVTFADGNGTTAEVTKRANGDSITVKYNKVVADGLKL 1688

RESULT 14
US-08-913-942-4
Sequence 4, Application US/08913942
Patent No. 6200578
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-942-4

Query Match 6.7%; Score 155.5; DB 4; Length 2353;
Best Local Similarity 19.7%; Pred. No. 0.0011;
Matches 103; Conservative 61; Mismatches 168; Indels 191; Gaps 22;

QY 70 AGGDDTTGVNAGGLNGRKGRTAGTTPOSDSQ-----NMLSEMNGNGLDQAITPDG 120
DB 1210 SAGNKEITNVKSA--LKYTKDONTADETQDKEFHAAVKNAVEFVGKNGATVSAKTDN 1267
QY 121 QG-----GGQIGDNPLLKAMKLIAARMMDGQSDQFGQPGTGN-----SASSGTSSS 167
DB 1268 NGKHTVTIDVAEAKVGDG-----LEKDTGKIKLVNDTGNLLTVDATKGASVA 1318
QY 168 GGSFFNDLGGKAPSGNSPVSTPSPPTSPSPSPPLDPPSPPKAAGSTPV-T 226
DB 1319 -----KGEFNAVTT---DATTAQGTNANERGVVYVKGSGNGATATET 1356
QY 227 DHPD--PVSAGIAGNSVAF-----TSAGANTV-LHDTITVKAGOVF- 267
DB 1357 DKKKVATGVDVAKAINDAATFVKVENDSDSATIDSDTDGANDALKAGDTLLKAGKNLK 1416
QY 268 ---DGKGQTFAGSE-----LGDGQSENQKPLFLEDCASL---KNVTMGD 308
DB 1417 VKRDGKNITFALANDLSVKVSATVSKLSLGTNGKNVN-----ITSOTKGLNFAKDSKTGD 1471
QY 309 DGADGTHLYG-----DAKIDNLHVTNVGEDAIT-----VKPN 340
DB 1472 DA--NIHLNGIASTLTDTLNLSGATTNLGGNGITDNEKKRAASVKOVNLNAGWNVGVKPA 1529
QY 341 SACKKSHVEITNSSFEHASKILQLNADTNLSVDNPKAKDFG-----WDLNLSHISAEDGKF 409
DB 1530 SANN---QVENIDFVATVDTVDVSGDKDTSVTVESKDNKGRTEVKIGAKTSVTKDHN 1585
QY 383 -----TFVRTNGGQGN-----WDLNLSHISAEDGKF 409
DB 1586 GKLTGKELKLDANNNGVTVTETDCKDEGNGLVTAKAVIDAVNAGWNVKVTGTGANGQNDDF 1645

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2001, 09:08:46 ; Search time 24.08 Seconds
(without alignments)
2455.994 Million cell updates/sec

Title: US-09-596-958-2

Perfect score: 2310

Sequence: 1 MSITLNNNTSSSGLFQSG.....LGDVENHYKVPMSANLKVAE 447

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_unclassified.*
- 13: sp_vertebrate.*
- 14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2310	100.0	447	054508	Q54508 erwinia amy
2	2277	98.6	447	09LAW2	Q9law2 erwinia amy
3	567.5	24.6	441	087327	O87327 pseudomonas
4	559	24.2	424	087264	O87264 pseudomonas
5	362.5	15.7	224	09RRW0	Q9rrw0 bacillus sp
6	353.5	15.3	221	034310	O34310 bacillus su
7	349.5	15.1	222	09X622	Q9x622 bacillus sp
8	226.5	9.8	266	09EX16	Q9ex16 streptomyce
9	216.5	9.4	215	0300843	O00843 nectria hae
10	207	9.0	1660	053215	O53215 mycobacteri
11	199.5	8.6	1306	053775	O53775 mycobacteri
12	194.5	8.4	233	0300845	O00845 nectria hae
13	193	8.4	240	093877	O93877 fusarium ox
14	193	8.4	694	053212	O53212 mycobacteri
15	191	8.3	929	09NBW0	O9nbw0 drosophila
16	191	8.3	929	09NBL3	O9nbl3 drosophila
17	191	8.3	939	09NHQ0	O9nhq0 drosophila
18	191	8.3	2280	09V8E6	Q9v8e6 drosophila
19	191	8.3	2302	09N693	Q9n693 drosophila

20	188	8.1	639	2	053952	053952 mycobacteri
21	186.5	8.1	591	2	007224	007224 mycobacteri
22	186	8.1	929	5	09NGW5	09ngw5 drosophila
23	186	8.1	1329	2	006810	006810 mycobacteri
24	185.5	8.0	809	5	P90534	P90534 dictyosteli
25	185.5	8.0	1381	2	053552	053552 mycobacteri
26	185.5	8.0	1489	2	053559	053559 mycobacteri
27	185	8.0	2310	5	09GRA9	09gra9 drosophila
28	183.5	7.9	242	3	004701	004701 fusarium so
29	183	7.9	588	2	050396	050396 mycobacteri
30	182.5	7.9	244	3	000851	000851 nectria hae
31	182	7.9	606	2	053884	053884 mycobacteri
32	182	7.9	731	2	050415	050415 mycobacteri
33	181.5	7.9	484	2	053394	053394 mycobacteri
34	180.5	7.8	1079	2	053557	053557 mycobacteri
35	179.5	7.8	588	14	Q9QEK6	Q9qek6 cynomolgus
36	179	7.7	730	10	Q9ZU23	Q9zu23 arabidopsis
37	179	7.7	1538	2	053395	053395 mycobacteri
38	179	7.7	3016	2	P73590	P73590 synechocyst
39	178	7.7	741	2	006808	006808 mycobacteri
40	177.5	7.7	2232	5	P91365	P91365 caenorhabdi
41	176.5	7.6	562	2	050458	050458 mycobacteri
42	176.5	7.6	2090	5	Q9W2T1	Q9w2t1 drosophila
43	175	7.6	645	2	053818	053818 mycobacteri
44	174	7.5	615	2	005806	005806 mycobacteri
45	174	7.5	2586	5	Q9VTK8	Q9vtk8 drosophila

ALIGNMENTS

RESULT	1
054508	
ID	054508
AC	054508; PRELIMINARY; PRT: 447 AA.
DT	01-JUN-1998 (TRENBLrel. 06, Created)
DT	01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT	01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE	HRPW PROTEIN.
GN	HRPW.
OS	Erwinia amylovora.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX	Erwinia.
OC	NCBI_TaxID=552;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CFBP1430;
RX	MEDLINE=98086111; PubMed=9426142;
RA	Gaudriault S., Malandrin L., Paulin J.P., Barny M.A.;
RT	"DspA, an essential pathogenicity factor of Erwinia amylovora showing
RT	homology with AvrE of Pseudomonas syringae, is secreted via the Hrp
RT	secretion pathway in a DspB-dependent way.";
RL	Mol. Microbiol. 26:1057-1069(1997).
RN	[2]
RP	SEQUENCE OF 138-447 FROM N.A.
RC	STRAIN=CFBP1430;
RA	Gaudriault S., Brisset M.N., Barny M.A.;
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE OF 1-138 FROM N.A.
RC	STRAIN=EA321;
RX	MEDLINE=98115919; PubMed=9448330;
RA	Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA	Conlin A.K., Collmer A., Beer S.V.;
RT	"Homology and functional similarity of an hrp-linked pathogenicity
RT	locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
RT	Pseudomonas syringae pathovar tomato.";
RT	Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RN	[4]
RP	SEQUENCE FROM N.A.
RC	STRAIN=EA321, ATCC 49947;
RA	Kim J.F., Zumoff C.H., Beer S.V.;
RT	"HrpW, a new harpin of Erwinia amylovora, is a member of a family of

Db 67 QNPTDSSAATDPOSNNVKKLLSALVTSLLQMLNKKQDTGQDSNEMQDPFQNEGGLG-- 124
 QY 89 KGTAGTTPOSNSMLSEMNGNGLDQAITPDGCGGQIGDNPPLKAMKLIARMMDGQSD 148
 Db 125 -----TPSA-----EGSDGT-----QEA 138
 QY 149 QFQPGTGNNSSAGTSSSGGSPFNDLGGKAPSGNSPS-----GNYSPVST-----FSPPS 200
 Db 139 SGDEGGGTAAATGDDGGGTSTTTEGDDG-----GTSPTEAGDGGGYSYSTGADGSGAPS 194
 QY 201 TPTSPSPLDPFSSPTKAAAGSTPVYDHPD--VGSAGIGAGNSVATTSAGANQ-----TV 254
 Db 195 T-----EDGTGGGGGSGVTPQVTPQLANPGRNSGNGTSDTTGSLQSGSEVNV 243
 QY 255 LHDTITVKGAVDFGKGTFTAGSELGDSGQSENOKPLFILEDGASLKNNVTMGDDGADGI 314
 Db 244 VRDTIKVGAGVDFGHGATFTADKSMGTGDDQDEHQPLFELAEAGVILKNVNGENEGADGI 303
 QY 315 HLYGDAK-----IDNLHVTNVGEDAITVKPNSAGKSKSHVEITNSFEHASDKILOLNAD 368
 Db 304 HV--NAKNSQVITIDNVHAQNVGEMDITVKEGGAKVTNLITNITNSANGADDKVIOLNAD 361
 QY 369 TNLSDNVKAKDGTGTVRNGGQ--GNWDLNLHSHISAEDGKSFVKSDEGLNWNVTSDIS 427
 Db 362 THLKVDGFRATDGTGLVRTNGGKQFDDMSVELNGVDATHGKFALVKSDSDDLKATGDIA 421
 QY 428 LGDVENHY 435
 Db 422 MTDVKHAY 429

RESULT 4
 ID O87264 PRELIMINARY; PRT; 424 AA.
 AC O87264;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE HRPW.
 GN HRPW.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=98422476; PubMed=9748456;
 RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
 RA Collmer A.;
 RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
 RT to harpins and pectate lyases and can elicit the plant hypersensitive
 RT response and bind to pectate."
 RL J. Bacteriol. 180:5211-5217(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=20243785; PubMed=10781092;
 RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
 RA Petnicki-Ocwieja T., van Dijk K., Collmer A.;
 RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
 RT mosaic structure composed of a cluster of type III secretion genes
 RT bounded by exchangeable effector and conserved effector loci that
 RT contribute to parasitic fitness and pathogenicity in plants."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
 DR EMBL: AF005221; AAC62526.1; -;
 DR EMBL: AF232006; AAF71503.1; -;
 SQ SEQUENCE 424 AA; 42910 MW; 7AF4ED059BE79D2E CRC64;

Query Match 24.2%; Score 559; DB 2; Length 424;
 Best Local Similarity 36.1%; Pred. No. 1.1e-27;

Matches 147; Conservative 54; Mismatches 132; Indels 74; Gaps 13;
 QY 37 QPTDROTISMAQLLAELLSL---LSPQSGNAATGAGNDQTTGYNAGGLNKRKGTAG 93
 Db 72 KPNDOS--NIAKLISALIMSLQMLTNSNKKQDTNOEOPDSOAPFONNGGLG----- 122
 QY 94 TTPQSDSNMLSMGNNGLDQAITPDGCGGQIGDNPPLKAMKLIARMMDGSDQFGQP 153
 Db 123 -TPSADS-----GGG-----TPDATGGG-GDTP-----SATGGG 151
 QY 154 GTGNNSASSGSSSGSPFNDLSSGRAPSGNSGNSVSPVSTPTSPSTPSPLDFFS 213
 Db 152 GGDTPATGGGSGGGGTPTATGG---SGGTPTATGGGEGGVTPTTQITPOL-----A 200
 QY 214 SPTKAAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANOTVLHDTITVKGAVDFGKGT 273
 Db 201 NPRTSG-----TGSVSDTAGS-----TEQAGKINVVKDTIKVGAAGEVDFGHGAT 245
 QY 274 FTAGSELGDSGQSENOKPLFILEDGASLKNNVTMGDDGADGIHLYG-----DAKIDNLHVTN 329
 Db 246 FTADKSMGNGDQGENQKPMFELAEAGATLKNVNLGENEVDGIHVAKAKNAQEVITDNVHAQN 305
 QY 330 VGEDAITVPKNSAGKSKSHVEITNSFEHASDKILOLNADTNLSVDNVKAKDGTGTVRNG 369
 Db 306 VGEDLITVKGEGGAATVNLNINIKNSAKGADKYVQLNANTHLKIDNFKADDFGTMVRTNG 365
 QY 390 GQO--GNWDLNLHSHISAEDGKSFVKSDEGLNWNVTSDISLGSDVENHY 435
 Db 366 GKQFDDMSIELNGIEANHGKFAVVKSDSDDLKATGNIAMTDVKHAY 412

RESULT 5
 ID O9RHWO PRELIMINARY; PRT; 224 AA.
 AC O9RHWO;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE PECTATE LYASE.
 OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1409;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KSM-P15;
 RA Hatada Y., Koike K., Salto K., Kobayashi T., Susumu I.;
 RT "Amino acid sequence and possible catalytic residues of a novel
 RT alkaline pectate lyase from alkaliphilic Bacillus."
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB011839; BAA87892.1; -;
 KW Lyase.
 SQ SEQUENCE 224 AA; 23751 MW; 99D04821B09DE523 CRC64;

Query Match 15.7%; Score 362.5; DB 2; Length 224;
 Best Local Similarity 46.1%; Pred. No. 1.1e-15;
 Matches 83; Conservative 29; Mismatches 59; Indels 9; Gaps 4;
 QY 248 AGANQTVLHDTITVKGAVDFGKGTFTAG--SELGDSGQSENOKPLFILEDGASLKNNVTM 306
 Db 25 AEAAPTVWHETIRVAPQGTGDKGTQVYVNPNTLGDGQSAENQKPIFRLEAGASLKNNVI 84
 QY 307 GDDGAGIHLGYDAKIDNLHVTNVGEDAITVKPNSAGKSKSHVEITNSFEHASDKILOLN 366
 Db 85 GAPAALSDGHCYGDCTITNVIWEDGADALTLK--SSGT---VNISGGAAYKAYDKVQFIN 139
 QY 367 ADTNLSVDNVKAKDGTGTVRNGGQGNWDLNLHSHISAEDGKSFVKSDE---GLNVNT 423
 Db 140 AAGTINIRFRADDIKLVQRNGGTTYKVVNVNENCNISRVKAILKIDTSSTSTGRIVNT 199
 RESULT 6

```

034310
ID   034310          PRELIMINARY;          PRT;      221 AA.
AC   034310;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YVPA.
DE   YVPA.
GN   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RP   Lazarevic V., Soldo B., Rivolta C., Reynolds S., Mauel C.,
RL   Karamata D.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP   SEQUENCE FROM N.A.
RP   STRAIN=168;
RC   MEDLINE=98044033; PubMed=9384377;
RX   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Chol S.K., Codani J.J., Connerthon I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis."
RN   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR   EMBL: AF017113; AAC67291.1;
DR   EMBL: Z99121; CAB15500.1;
SQ   SEQUENCE 221 AA, 24281 MW:  AB324700DE573236 CRC64;

```

			Query Match	15.3%	Score 353.5;	DB 2;	Length 221;	
			Best Local Similarity	41.8%;	Pred. No. 4e-15;			
			Matches	76;	Conservative	31;	Mismatches	70;
							Indels	5;
			Gaps					1;
QY	250	ANQTVLHDTTVVAGQVFDSKGTFTAGSELGGGQSENQKPLFILEDGASLKNVTWGDD	309					
		: ::	:				:::	
Db	27	AADKVWHETIIVPKNTTYDGQRFAVGKELGDCGSQENQDPVERVEDGATLKNNVLGAP	86					
			:				:::	
QY	310	GAOGIHLYGDAKDINLHVTVNGEDAITVKPNSACKKSHVEITNSFPHASKILQLLNADT	369					
		: ::	:				:::	
Db	87	AACGVHTHYGNVNTQNKWVEDGDALTVK-----KEGKVTIDGSGAKSKAQSDIFQINKAS	141					
			:				:::	

Qy	370	NLSVDNVKAKDFGTFVVRTNGCGQGGNNDLNLSHISAEDGKSPVKSDSEGLNVTNSDISLG	429
Db	142	TFTVKNFTADNGGKFRQLGGSFTFHDVLDIKDTTINMKEAIFRTDSKSTSTVRMTNTRY	501
Qy	430	DV 431	
Db	202	NV 203	
RESULT	7		
Q9X6Z2		PRELIMINARY; PRT; 222 AA.	
ID	Q9X6Z2		
OC	Q9X6Z2;		
DT	01-NOV-1999 (TrEMBLrel. 12, Created)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)		
DE	PECTATE LYASE PRECURSOR (EC 4.2.2.2).		
GN	PELA.		
OS	Bacillus sp. BP-23.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=89769;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BP-23;		
RA	Sortiano M., Blanco A., Diaz P., Pastor F.I.J.;		
RL	Submitted (APR-1999) to the EMBL/GenBank/DBSJ databases.		
DR	EMBL; A0237980; CAB40884.1; -.		
KW	Signal; Lyase.		
FT	SIGNAL	1	25
CHAIN	26	222	PECTATE LYASE.
SEQUENCE	222 AA, 23233 MW;		
SEQ	7F29F585791C9682		
CR	CRC64;		

[illegible]

RESULT	8
Q9EX16	
ID	PRELIMINARY; PRT; 266 AA.
AC	Q9EX16;
DT	01-NAR-2001 (TremBurel. 16, Created)
DT	01-NAR-2001 (TremBurel. 16, Last sequence update)
DT	01-NAR-2001 (TremBurel. 16, Last annotation update)
DE	POTATIVE SECRETED LYASE.
DE	2SCG38.03.
OS	Streptomyces coelicolor.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces
OX	NCBI_TaxID=1902;
OX	[1]
RP	SEQUENCE FROM N.A.
RP	STRAIN=A3(2);
RA	Saunders D.C., Hartsis D.;


```
Db 634 GCGAGGCGDPCAGCKGCGAG-----GAGATEGVTGATGATVHSG--G 674
QY 234 SAGIGAGSVAFSTAGANTVLHDTITTVKAGQVFDKGGTFTAGSELGDDGGQSENQKPLF 293
Db 675 NGKKG-GNCADATVAGN-----CGKGGAGNGGLVGDGGAG----- 710
QY 294 ILEDG-----ASLKNVTMGDDGADG 313
Db 711 --GDGSGAAGANGANGVEDGADG 732

RESULT 11
ID O53775 PRELIMINARY; PRT; 1306 AA.
AC O53775;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PGRS-FAMILY PROTEIN.
GN RV0578C OR MV039.16C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA - Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; AL021942; CAAL7449.1; -.
DR TuberculList; Rv0578C; -.
DR InterPro; IPR000084; -.
DR InterPro; IPR002029; -.
DR InterPro; IPR002173; -.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; -. 1.
DR PROSITE; PS00583; PKB_KINASES.1; UNKNOWN.2.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN.1.
SQ SEQUENCE 1306 AA; 105964 MW; 843A30955FFA56B6 CRC64;
```

```
Query Match 8.6%; Score 199.5; DB 2; Length 1306;
Best Local Similarity 26.0%; Pred. No. 0.00014;
Matches 88; Conservative 16; Mismatches 122; Indels 113; Gaps 11;

QY 8 NWTSSSPGLFQSGDNGLGHHNANSALGOQPIDROTIEQMAQLLELLKSLSPQSGNAA 67
Db 754 NGVAGSQGGAGGGTGGTGGVGGNGRGIDGADGAT----- 789
QY 68 TCAGGNDQTTGVNAGNLNKGKTAGTTTQSDSNQNLSEMGNNGLDQAITPDGCGGGQIG 127
Db 790 --AGARGDGGAGGAGGKGGRGTTGP-----GGAGPAGTTGSGQAGGNG 832
QY 128 DNPLKAMLKLIARMMDGOSDQFGOPGTGNNSASSTSSGSPFNDLSGGKA-PSGNSP 186
Db 833 -----GSGGTGGDPDGGNGANGSVFTNNGIGGNGGNGAGPSGAGG 875
QY 187 SGNYSVPSTFSPPTPTSPSL-----DFPSSPTKAG-----G 221
Db 876 SGGAG--STFG----ATGSSSIHVNGGNGGNGDHALSGAAGGNGGNGSLRG 929
QY 222 STPVTDHPDPCVSGAGIGAGNSVAFISAGANQIVLHDTITTVKAGQVFDKGGTFTAGSELG 281
```

```
Db 930 SGGAGGHHGNGGNGARGMGGDGGTGGAGGN-----AGQICNG-----GAGNGG 973
QY 282 DGGQSENQKPLFIL-----EDGASLKNVTMGDDGADG 313
Db 974 DGGTSGDGNPAGITGSGRGGDGGVGGGGGSGVAGDGDG 1012

RESULT 12
Q00845 PRELIMINARY; PRT; 233 AA.
AC Q00845;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PECTATE LYASE D.
GN PELD.
OS Nectria haematococca mpVI.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Nectria.
OX NCBI_TaxID=70791;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=T8;
RX MEDLINE=96400363; PubMed=8806739;
RA Guo W., Gonzalez-Candelas L., Kolattukudy P.E.;
RT "Identification of a novel pelD gene expressed uniquely in planta by
RT Fusarium solani f. sp. pisi (Nectria haematococca, mating type VI) and
RT characterization of its protein product as an endo-pectate lyase.";
RL Arch. Biochem. Biophys. 332:305-312(1996).
DR EMBL; U13050; AAC49420.1; -.
SQ SEQUENCE 233 AA; 24509 MW; 2FF53A0D9A4E2CB9 CRC64;

Query Match 8.4%; Score 194.5; DB 3; Length 233;
Best Local Similarity 29.9%; Pred. No. 3.9e-05;
Matches 50; Conservative 34; Mismatches 74; Indels 9; Gaps 3;

QY 260 TVKAGQVFDGKGQTFTAGSELGDDGGQSENQKPLFLEBDGASLKNVTMGDDGADGHLXYCD 319
Db 40 TLKKGVEFDAGWVRVDRGVKCSQAEQSGKDAVFILEEGATLRNVIIGANQREGIHCXGS 99
QY 320 AKIDNLHVTNVEDAITVKPNASAGKSHVETNTSFEHASDKILOLNADTNLSVDNVKAK 379
Db 100 CNIEFAPWEDVEDAIISILSGT-----ANIIGGAYHASDKVIOHNGCGHVNIVFYAN 154
QY 380 DFGTFVRTNGGQGNWDLNLS-HISAEDGKFSFVKSDSEGLNVTSD 425
Db 155 DYKGYRSCGNCKNGTNCCKRSVHM---EGTAVKGGELIGINTNYGD 198

RESULT 13
Q93877 PRELIMINARY; PRT; 240 AA.
AC Q93877;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PECTATE LYASE.
OS Fusarium oxysporum f. sp. lycopersici.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Fusarium.
OX NCBI_TaxID=59765;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=42-87;
RA Huertas-Gonzalez M., Ruiz-Roldan M., Garcia-Maceira F., Roncero M.,
RA Di Pietro A.;
RT "Cloning and characterization of pII encoding an in planta expressed
RT pectate lyase of Fusarium oxysporum.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF080485; AAC64368.1; -.
KW Lyase.
```

[illegible]

Search completed: October 18, 2001, 09:10:32
Job time: 106 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2001, 09:08:46 ; Search time 13.15 Seconds
(without alignments)
1164.426 Million cell updates/sec

Title: US-09-596-958-2

Sequence: 1 MSILTLNNNTSSSGLFQSG.....LGDVENHYKVPMSANLKVAE 447

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	177.5	7.7	1901	1 YZ08_MYCTU
2	171	7.4	398	1 CSP_PLACC
3	171	7.4	463	1 Y468_MYCTU
4	169.5	7.3	914	1 WA22_MYCTU
5	166	7.2	603	1 YD25_MYCTU
6	165	7.1	959	1 N100_YEAST
7	163	7.1	400	1 RTOA_DICDI
8	162	7.0	434	1 YK98_MYCTU
9	160.5	6.9	314	1 YR83_MYCTU
10	160.5	6.9	937	1 HYR1_CANAL
11	160.5	6.9	957	1 Y278_MYCTU
12	159.5	6.9	778	1 YQ34_MYCTU
13	157.5	6.8	2329	1 YS89_CAEEL
14	157	6.8	543	1 YP91_MYCTU
15	155	6.7	1113	1 N116_YEAST
16	155	6.7	2249	1 I90K_RICRI
17	154	6.7	403	1 HRPN_ERWAM
18	154	6.7	430	1 NU42_YEAST
19	153	6.6	788	1 SPI_RAT
20	153	6.6	1049	1 CAL3_BOVIN
21	152.5	6.6	2003	1 YDBA_ECOLI
22	150.5	6.5	598	1 YAU8_SCHPO
23	150	6.5	515	1 Y140_MYCTU
24	148	6.4	801	1 Y747_MYCTU
25	147	6.4	1453	1 CAL1_CHICK
26	146.5	6.3	366	1 CAS4_EPHMU
27	146	6.3	937	1 NU98_RAT
28	145	6.3	1736	1 CA2B_HUMAN
29	144	6.2	1156	1 GLH4_CAEEL
30	144	6.2	1581	1 PPRB_HUMAN
31	141.5	6.1	419	1 CSP_PLACC
32	141.5	6.1	541	1 NU57_YEAST
33	141	6.1	1150	1 APMU_PIG

34 140.5 6.1 2132 1 PGCA_MOUSE
35 140 6.1 671 1 HMOC_DROME
36 140 6.1 678 1 YF48_MYCTU
37 139 6.0 1325 1 YDEK_ECOLI
38 138.5 6.0 630 1 MUC1_MOUSE
39 138.5 6.0 1475 1 N153_HUMAN
40 138.5 6.0 1516 1 NCO2_XENLA
41 138 6.0 671 1 CAL1_RAT
42 138 6.0 2090 1 N214_HUMAN
43 138 6.0 3591 1 FHAB_BORPE
44 137.5 6.0 283 1 YQ33_CAEEL
45 137.5 6.0 573 1 C114_MOUSE

ALIGNMENTS

RESULT 1
YZ08_MYCTU
ID YZ08_MYCTU STANDARD; PRT; 1901 AA.
AC O53553:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN RV3508 OR MTV023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
Taylor K., Whitehead S., Barrett B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AL022022; CAA17745.1; -;
DR HSP; P19972; 1KVD.
DR TuberculList; RV3508; -;
DR InterPro; IPR000084; -;
DR Pfam; PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
RV3508.
FT SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match 7.7%; Score 177.5; DB 1; Length 1901;

Best Local Similarity 25.9%; Pred. No. 0.049;

Matches 79; Conservative 22; Mismatches 109; Indels 95; Gaps 11;

QY 20 GGDNGLGHGHNANSALGQPIDRQTIETQMAOLLAEKLSLSPQSNATGAGNDQTTGV 79

Db 599 GSSGAGGTNGSGGAGG-----TGCGGAGGAGGAGADNPCTGI 636
QY 80 GNAGGLNRKRGTAGTTPQSDSNMLSEMNGLDQAITPDGCG--QOIGDNPLLKAMLK 137
Db 637 GGAGGTGTTGGAAGA-----GAGGAIGTGTGGGAVGVGNAGIGGT--- 678
QY 138 LIARMMDGSDQFGQGTGNNSSASGTSSSGSPFNDSLGGK-APSGNSPSGNSYSPVSTF 196
Db 679 -----GCTGVGAGGAGAGAAAGSSATGAGFAGGAGGEGGAGGNSGVG----- 723
QY 197 SPSPTPTSPPLDPSPPTKAAGSTPVTHDPVGSAG-----IGAGNSVAFVTS 248
Db 724 ----GTNGSGGAGGAGGKGTGAGGSG--ADNPTGAGFAGGAGGTGGAAGGAGGATGT 778
QY 249 GANQTVLHDTITVKAGQVDFGKGTFTAGSELGDSQSENQKPLFILEDGASLKNVTWGD 308
Db 779 GTGGVVGAT-----GSAGIGGAGRGDGG-----DGAS--GLGLGL 814
QY 309 DGADG 313
Db 815 SGFDG 819

RESULT 2

CSP_PLACC STANDARD; PRT; 398 AA.
AC P08673;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS Plasmodium cynomolgi (strain Ceylon).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5829;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87102878; PubMed=3802196;
RA Galinski M.R., Arnot D.E., Cochran A.H., Barnwell J.W.,
RA Nussenzweig R.S., Enea V.;
RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RL Cell 48:311-319(1987).
CC -!- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M15103; AAA29533.1; -.
DR PIR; C26255; OZQAS.
DR InterPro; IPR000884; -.
DR InterPro; IPR003067; -.
DR Pfam; PF00090; tsp_1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
KW Malaria; Sporozoite; Repeat; Signal.
FT SIGNAL 1 19 PROBABLE.
FT CHAIN 20 398 CIRCUMSPOROZOITE PROTEIN.
FT DOMAIN 97 249 17 X 9 AA TANDEM REPEATS OF A-G-N-N-A-A-
FT A-G-E.
SQ SEQUENCE 398 AA; 37718 MW; 60FA2E8A62ED05BF CRC64;

Query Match 7.4%; Score 171; DB 1; Length 398;
Best Local Similarity 24.4%; Pred. No. 0.019;
Matches 89; Conservative 46; Mismatches 162; Indels 68; Gaps 15;
QY 4 LTLNNTSSCLFQ-----SGDNGLGHNNSALGQQPIDRQTTEQMAQLLAEKLSLL 59
Db 37 VSFNVDAASLGAQVROSASRGRLGENPKNEEGADKP--KKDEKOVEPKPKRENK 94
QY 60 SPQSNAAATGAGNDQTTG-VGN---AG--GLNGRKGTAGTTP---QSDSNMLSEMNN 110
Db 95 QPAGNNAAGEAGNNAAGEAGNNAAGEAGNNAAGEAGNNAAGEAGNNAAGEAGN 154
QY 111 GLDQAITPDGGGGIGDGNPLKAMLKLIARMMDGSDQFGQPG-----TGNSAS-- 161
Db 155 AAGGA-AGNNAAGEAGN-----AAGGAAGNNAAGEAGNNAAGEAGNNAAGE 204
QY 162 SCTSSSGSGSPNDLSGGKAPSGNSPSGNSYSPVSTPTSPPLDFFSPPTKAAGG 221
Db 205 AGNNAAGGAAGNNAAGEA--GNNAAG-----AAGNNAAG 239
QY 222 STPVTDHPDPVGSAGIG-AGNSVAFTSAGANQTVLHDTITVKAGQVDFGKGTFTAGSEL 280
Db 240 AAGNNAAGEAGAGAGAGAGNNAAGEAGAGAGAGAGAGAGAGAGAGAGAGAGAG 298
QY 281 GDGQSENQKPLFILEDGASLKNVTWGDGAGIHYGDAKIDNLHVTNVGDAITVVKPN 340
Db 299 GDAGGQGN-----NGGANVPNVKLVKEYLDKIRSTIGVEMSPGCVTCGKGVRRRKVN 351
QY 341 SAGKK 345
Db 352 AANKK 356

RESULT 3

YA68_MYCTU STANDARD; PRT; 463 AA.
ID YA68_MYCTU
AC O53416;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV1068C.
GN RV1068C OR MV017.21C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL021897; CAA17184.1; -.

SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z73902; CAA98089.1; -;
 DR HSSP; P1972; IKVD.
 DR Tuberculin; Rv1325C; -;
 DR InterPro; IPR000084; -;
 DR Pfam; PF00934; PE; 1.
 KW Hypothetical protein; Repeat; Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 603 HYPOTHETICAL PE-PCRS FAMILY PROTEIN
 FT 114 603 Rv1325C.
 FT DOMAIN 114 603 GLY-RICH.
 FT SEQUENCE 603 AA; 49575 MW; 4F9BCB82B07AE964 CRC64;
 SQ -----

Query Match 7.28; Score 166; DB 1; Length 603;
 Best Local Similarity 24.98; Pred. No. 0.034;
 Matches 83; Conservative 15; Mismatches 130; Indels 106; Gaps 12;

Qy 10 TSSSPGLFQSGDNLGCHNANSALGQOPIDRQTTEOMAQLLAELLKSLSPQSGNAATG 69
 Db 274 TGGGGGLFSGAGGAGGFGVSGSAG-----GNGGTG 305

Qy 70 A-----GNDQD--TCVGN--AGLNGRKGTAGTTTPQSDSQNMLSEMGNGLDQAIT 117
 Db 306 GDGIGFTGNGGTGGTGTGNLQVGGEGGAGGAGNAGILFGAGGIGTGTGLG-APD 364

Qy 118 PDQGG-QQIGNPLLKAMKLKLIARMMDQSDQFCQPGTGNNSASSGSSGGSPFNDS 176
 Db 365 PGGTGGKGGVGG-----IGGAGALFGPGAG-----GTGGFGASSADQMA 404

Qy 177 GKKAPSGNS-----PSGNYSPVSTFSPPTPTPTSPPLDFPSPPTKAAGGSTPV 225
 Db 405 GGIGSGSGGGAALIGDGGAGGTGDSVRGAGSGGTGGTGL-----IGDGGAGGAGGT 460

Qy 226 TDHPDPVSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVDFDKGQFTTAGSELGDDGG 285
 Db 461 GIEFGSVGGAGGAGGNAAGLSCAG-----GAGGAGGFGETAGDGA 501

Qy 286 SENQKPLFILEDGASLKNVTMGDDGADGHIHYGD 319
 Db 502 CGN-----AGLLNGDGGAGGAGGIGTAGD 525

RESULT 6
 ID N100_YEAST STANDARD; PRT; 959 AA.
 AC Q02629;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE NUCLEOPORIN NUP100/NSP100 (NUCLEAR PORE PROTEIN NUP100/NSP100).
 GN NUP100 OR NSP100 OR YKL068W OR YKL336.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054906; PubMed=1385442;
 RA Wente S.R., Rout M.P., Blobel G.;
 RT "A new family of yeast nuclear pore complex proteins.";
 RL J. Cell Biol. 119:705-723(1992).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=S288C;
 RX MEDLINE=94378724; PubMed=8091863;
 RA Rasmussen S.W.;
 RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
 RT NUP100 gene, an open reading frame (ORF) possibly representing a
 RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
 RT addition to seven ORFs with weak or no significant similarity to
 RT known proteins.";
 RL Yeast 10:S69-S74(1994).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLOCATION OF
 CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
 CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z15035; CAA78753.1; -;
 DR EMBL; X75780; CAA53406.1; -;
 DR EMBL; Z28068; CAA81905.1; -;
 DR PIR; B44402; B44402.
 DR PIR; S39173; S39173.
 DR PIR; S44518; S44518.
 DR SGD; S0001551; NUP100.
 KW Nuclear protein; Transport; Repeat.
 FT DOMAIN 33 571 G-L-F-G.
 FT SEQUENCE 959 AA; 99988 MW; D3985F9901BBAA51 CRC64;
 SQ -----

Query Match 7.18; Score 165; DB 1; Length 959;
 Best Local Similarity 22.98; Pred. No. 0.1;
 Matches 117; Conservative 51; Mismatches 198; Indels 146; Gaps 23;

Qy 5 TLNNTSSPGLF-----QSGDNLGCHNANSALGQOPIDRQTTEOMAQLLAELLKS 57
 Db 261 TLGNTNNRNLGFGOMSSNGSGSLFGQNSMNS-STQGVFGNNQM-QINGNNNS 318

Qy 58 LLSPQS--GNAATGA-----GNDQDTGVGNAGLNGRKGTAGTTTPQSDSQ 101
 Db 319 LFGRANTFSNASGSLFGQNNQQSGSLFGQNSQTSQSGSLFGQNNQK-QPNTTQSN TG 377

Qy 102 NMLSEMGNGLDQA-----ITPDGQGGQIQDNPLLKAMKLKLIARMMDQSDQFGQP--- 153
 Db 378 IGLFGQNNNQOQSTGLFGAKPAGTGTGSLFGN-----SSTQPNSL 418

Qy 154 -GTGNNSASGTSSTSSGSPFNDLGGKAPSGNSPGNSPVSTFSPPTPTPTSPPLDFP 212
 Db 419 FGTTNVTSTNTQSQGNSLFGATKLTNMPFGNPTANQSGNSLFGTKPASTTGSGL---- 475

Qy 213 SSPTKAAGGSTPVDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDKGQ 272
 Db 476 -----FGNNTASTVPTSTNGLFGNANNSTTTNTGL-----FGAKPDSQSK 517

Qy 273 TTAGSELGDCGQSEN-----OKPLFILEDGASLKNVTMGDDGADGHIH-----LYGD 319
 Db 518 PALGGGLFGNSNSNSTIGQNKPVF---GGTTQNT--GLFGATGTNSSAVGSTGKLF- 570

Qy 320 AKIDNLHVTNVGDAITVKPNSACK-----SHVEITNSSFEHASDKILQNLNADT 369
 Db 571 ---QNNNTLVGTQNVPPVNTTQALLGTTAVPSLQAPVTN---EQLFSKISIPNSIT 624

Qy 370 NLSVDNVKAKDGTFTVRTNGQQGQNMWLNLSH-----ISAEDGKFSFYK----- 413
 Db 625 N-----PVKA-----TTSKYNADMKNRNSLTSAYRLAPKPLFAPSSNGDAKFQKWLTLRS 676

QY 414 -----SDSEGLNVNTSISLGDVENHY 435
Db 677 DRGSSTNSITDPESSYLNSDL-LFDPDRY 707

RESULT 7
ID RTOA_DICDI STANDARD; PRT; 400 AA.
AC P54681;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE RTOA PROTEIN (RATIO-A).
GN RTOA.

OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97108762; PubMed=8951083;
RT Wood S.A., Ammann R.R., Brock D.A., Li L., Spann T., Gomer R.H.;
RT "RtoA links initial cell type choice to the cell cycle in
Dictyostelium".
RL Development 122:3677-3685(1996).

CC -!- FUNCTION: MAY HAVE DUAL FUNCTIONS, ONE IN VEGETATIVE CELLS AT
STARVATION INVOLVING CELL-TYPE CHOICE AND A LATER FUNCTION
INVOLVING PROGRESSION TO THE TIPPED MOUND STAGE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN VEGETATIVE CELLS, LEVELS
DECREASE DRAMATICALLY AS DEVELOPMENT BEGINS, AND REMAIN LOW
THROUGHOUT THE AGGREGATION AND LOOSE MOUND STAGES. THE LEVELS
BECOME MORE ABUNDANT AGAIN AT THE TIGHT-MOUND STAGE AND REMAIN
HIGH THROUGHOUT THE REST OF DEVELOPMENT AND FRUITING BODY
FORMATION.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

CC EMBL; U48298; AAC47311.1; -
DR DictyDb; DD02056; rtoA.
KW Transmembrane; Repeat.
FT TRANSMEM 30 50
FT DOMAIN 30 38
FT DOMAIN 87 96
FT DOMAIN 181 181
FT REPEAT 181 190
FT REPEAT 191 200
FT REPEAT 201 211
FT REPEAT 212 222
FT REPEAT 223 233
FT REPEAT 234 244
FT REPEAT 245 255
FT REPEAT 256 266
FT REPEAT 267 277
FT REPEAT 278 287
FT REPEAT 288 298
SQ SEQUENCE 400 AA; 39840 MW; 0392F4E68DC27A75 CRC64;

Query Match 7.1%; Score 163; DB 1; Length 400;
Best Local Similarity 21.8%; Pred. No. 0.049;
Matches 83; Conservative 54; Mismatches 151; Indels 92; Gaps 12;
QY 7 NNTSSSGPLFGGCGDGLGGHNSALCOQPIDRTIQMAQLAELLKLLSPQSGNA 66
Db 78 SNTASSEGVSSSSNSG-----SOSTSNGSGE 105
QY 67 ATGA--GGNDQTTGVGNAGLGRKGTAGTTPQSDSQNMLSEMNGLDQAITPDQGGG 124

Db 106 ASGSSNSGSGSTNSGSEASGSSNSGSGSTDSNSGSGSTGSSNSGSGSTDSNSGSGS 165
QY 125 QICDNPLLKAMLKLIARMMDGQSDQFGQGTGN--SASGTSSSGG--SPNDLSGKGKAPS 182
Db 166 QSTSD-----SSNSGSGSGSTGSSNSGSESSGSSGSSGSSGSSGSSGSS 211
QY 183 GNSPSGNSPVSTESPSTPTSPDPLDPSPPTKAAGSTPVTDHPDPVGSAGICAGNS 242
Db 212 GSSNSGSESSGSSNSGSGSSGSSNSGSESSGSSNSGSESSGSSNSGSESSGSSNS 271
QY 243 VAFTSAGANOTVLHDTITVKAGQVFDKGQTFTAGSELGQSGQSENQKPLFI--LEDGASL 301
Db 272 GSESSGSS-----NSGS--ESSGSSNSGSESSSDGSSSDGKTTCTISFHDLSI 319
QY 302 KNYVTMGDDGADGIHLGYDAKIDNLHVTNVGEDAITVKPN--SAGKSHVEITNSSFHAS 359
Db 320 NTV---DD-----DEICTGKTRCISDNNYKCATKORHGSIECS----- 357
QY 360 DKILQNLNADTNLSVDNVKAK 379
Db 358 -----VNGYIRCTGSNIKCK 372

RESULT 8
ID YK98_MYCTU STANDARD; PRT; 434 AA.
AC Q10707;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 36.5 KDA GLYCINE-RICH PROTEIN RV2098C.
GN RV2098C OR MTCV49_38C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGSS
SUBFAMILY.

CC -!- CAUTION: THIS PROTEIN IS FRAMESHIFTED NEAR THE N-TERMINUS.
THE SEQUENCE HAS BEEN CHECKED BY AUTHORS IN REF.1 AND THEY REPORT
THAT NO ERRORS HAVE BEEN FOUND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z73956; CAA98228.1; -
DR HSSP; P19972; 1KVD.
DR Tuberculist; RV2098C; -
DR InterPro; IPR000084; -
DR InterPro; IPR002952; -
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.

KW	Hypothetical protein; Transmembrane.
FT	TRANSMEM 23 43 POTENTIAL.
FT	TRANSMEM 98 118 POTENTIAL.
FT	TRANSMEM 221 241 POTENTIAL.
SO	SEQUENCE 314 AA; 30773 MW; 81E97B754ECED1E CRC64;
Query Match 6.9%; Score 160.5; DB 1; Length 314;	
Best Local Similarity 28.8%; Pred. No. 0.05;	
Matches 75; Conservative 25; Mismatches 71; Indels 89; Gaps	
QY	16 LFQSGDGNLGGHNANSALGQQPTDRGTIEQMAQLLAELKLSLPQSNA--TGA 70 : : : :
Db	42 LFGLTWNADLGGANAD-----IEGDDTRRAHAADAARK--FSANEANAQMOCVGA 92 : : : :
QY	71 GGNDQ-ITGVGNA--GGLNGRKCTAGTTPOSDSQNMLSEMGNNGLDQAITPDGOGGQGIG 127 : :
Db	93 QGMAQMASGITGALSGLGVMGPLTLQPQAMQ-----AGQAMQ-- 133 : :
QY	128 DNPLLKAM-----LKLI--ARMMDQSDQGPCTGNNASASSGTSSSGSPFNDLSG 177 : :
Db	134 --PLMSAQQAQAGDLAAVDGARLDLSIG---GEPSLG-----SGAG--GDVGG 177 : :
QY	178 GKAPSGNSPNSGNY--PVSTFPSPSTPTSPLDFESSPTKAA-----CGSTPVTDH-- 228 : :
Db	178 GGA-GGTTPTGYLGPPVPVTSSEPPTPA-----GAPTKSNTWPPPGASPASAHM 227 : :
QY	229 -----PDVGSAGIGAG 240 : :
Db	228 AAGPMVPPGMARGEGSG 247 : :

RESULT	10
HYRI_CANAL	
ID	HYRI_CANAL STANDARD; PRT; 937 AA.
AC	P46591;
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	HYPHALLY REGULATED PROTEIN PRECURSOR.
GN	HYRI.
OS	Candida albicans (Yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC	Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX	NCBI_TaxID=5476;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 10261;
RX	MEDLINE=96404781; PubMed=8808922;
RA	Bailey D.A., Feldmann P.J.F., Bovey M., Gow N.A.R., Brown A.J.P.;
RT	"The Candida albicans HYRI gene, which is activated in response to hyphal development, belongs to a gene family encoding yeast cell wall proteins.";
RL	J. Bacteriol. 178:5353-5360(1996).
CC	-I- FUNCTION: NONESSENTIAL COMPONENT OF THE HYPHAL CELL WALL (POTENTIAL).
CC	-I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).
CC	-I- DEVELOPMENTAL STAGE: ABUNDANT IN HYPHAE.
CC	-I- INDUCTION: INDUCED SPECIFICALLY IN RESPONSE TO HYPHAL DEVELOPMENT
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC	at the European Bioinformatics Institute. There are no restrictions on the
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commer-
CC	cial entities requires a license agreement (See http://www.isb-sib.ch/anno-
CC	or send an email to license@isb-sib.ch).
CC	-----
EMBL; Z50123; CAA030485.1; --	
KW Cell wall; Glycoprotein; Signal; GPI-anchor; Transmembrane; Repeat.	
DR SIGNAL 1 ? POTENTIAL.	
FT CHAIN ? 913 HYPHALLY REGULATED PROTEIN.	


```
Db 230 IFTMNVVLTPTTPTPTTSSDDAGGTGGTGATG----- 266
QY 59 LSPQSGNAATGAGNDQTTGVGNA-----GGLNGRKG-----AGTTPQSDSONMLSE 106
Db 267 -----GTGCTGGSGSATTLTSTGDAVRSTTSFGSGSGSSTGSGAGGSGTTASGSGSGSG 321
QY 107 MGNGLDQ-----AATPDGOG-----GQIGDNPLLKAMLKLIARMMDGQSDQFCQPT 155
Db 322 TGSQGVNSGKTTALNGDGTGSGTATTGSHLGD----- 354
QY 156 GNNASGTSSTSGGSPFN-DLGGKAPGNS-PSGNYSPVSTFSPSTPTSPPLDFFPS 213
Db 355 GSGTSGSGSDSGSGSVTKSSGSDTSSGSSGANGAFSAQAQPSPTTTRSSLATV 414
QY 214 SPTKAAGSTPTVTHDPDPYGS-AGGAGNSVAFTSAGANQTVLHDTITVKAGQVDFGKQ 272
Db 415 SPISAAEQAIIDAQADVNMQLAGIMDG-SASNNLSLNTSSLLNQISSLPAAADLVEVAQS 473
QY 273 TTAGSELGDGQSGENQKPLFILEDGASLKNTMGDDGADGHLGYDAKIDNLHVTNV-- 330
Db 474 LLSNLTLPKIPGVGNMSSVDVLKTLQDNIAATNSLADMAKVI-----TKLANYNMTSAQS 528
QY 331 -----GEDAIVKPSAGKSHVEITNSFEHAS-----DKIL 363
Db 529 LNSVLSLIDLALKGSTVYTLGVSSTKSKDGTVAIFGYVIASGYTLVSPRCTLSITVGSII 588
QY 364 QLNADTNLSV-----DNVAKDGTFTVRTNGGQGNMNLNLSHISAEDKGFVKSQDSEG 418
Db 589 YLTGDTFRASYKOLDGTVTA---DTMLAAAGTQGFATNGRTVQVEQDKIDDKRSLVSG 645
QY 419 LNVNTSDISLGDVEN 433
Db 646 -NIMATMSGVDVQVS 659

RESULT 14
Y991_MYCTU
ID YP91_MYCTU STANDARD; PRT; 543 AA.
AC Q50630;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEICAL PE-PCRS FAMILY PROTEIN RV2591.
GN RV2591 OR MTC227.10C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtrovd S.,
RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PCRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; 277724; CAB01283.1; -
DR TuberculList; RV2591; -
DR InterPro; IPR000084; -
DR Pfam; PF00934; PE; 1.
KW Hypothetical protein.
SQ SEQUENCE 543 AA; 46287 MW; 59730339ESD2DF59 CRC64;

Query Match 6.8%; Score 157; DB 1; Length 543;
Best Local Similarity 23.1%; Pred. No. 0.14;
Matches 81; Conservative 39; Mismatches 143; Indels 88; Gaps 14;

QY 1 MSTLTNNNTSSPGLFQSGDN-----GLGHNANSALGQPIDRQTTEQMAQLAE 54
Db 1 MSFVT-----AAPEMLATAAQNANIGTSLSAANATAAATTSVLAAGADEVSAIARL 54
QY 55 LKSLSP-OSGNAATGAGNDQTTGVNAGGLNGRKGTAAGTTPQSDSONMLSE----- 106
Db 55 FSDYATHYQSLNAAQAFAHHSFVOTLNAAGGAYSSAEANASQAQLEONLLAVINAPQA 114
QY 107 -----MGNNGLDQAITPDGQGGQI-----GDNPLLKAMLKLIARMMDGQSDQFCQPTGN 157
Db 115 LFCRPLIGNANGCTAASPNGDGGILYNGGN-----GFSQTTAGVAGGAGSGAGLNGG 170
QY 158 NSASSGTSSTSGGSPFNLDLGGKAP--SGNSPSGNYSPVSTFSPSTPTSPPLDFFSSP 215
Db 171 NGGAGGAGAAG-----AGGAGGWLGLNGGAG-----GPGGPTDVPAGT 209
QY 216 TKA--AGGSTPTVTHDPDPYGSAGIGA-GNSVAFTSAGANQTVLHDTITVKAGQVDFGKQ 272
Db 210 GGAGGAGGAPL-----IGWNGGPGGGAFFNGG-----AGNGGASGS 250
QY 273 TTAGSELGDGQSGEN-----QKPLFILEDGASLKNTMGDDGADG 313
Db 251 LFCVGGAGGVGSGSDVGTGTGAGGAGRGLFLGLGGDGGAGGTSNNNGDG 301

RESULT 15
N116_YEAST
ID N116_YEAST STANDARD; PRT; 1113 AA.
AC Q02630;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NUCLEOPORIN NUP116/NSP116 (NUCLEAR PORE PROTEIN NUP116/NSP116).
GN NUP116 OR NSP116 OR YMR047C OR YMR532.12C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins."
RL J. Cell Biol. 119:705-723(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93099880; PubMed=1464327;
RA Wimmer C., Doye V., Grandi P., Nehrass U., Hurt E.C.;
RT "A new subclass of nucleoporins that functionally interact with
RL nuclear pore protein NSP1."
RN EMBO J. 11:5051-5061(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBUNIT: INTERACTS WITH KAP95.
```

Job time: 127 sec

```
CC -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC -!- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -!- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15036; CAA78754.1; -
DR EMBL; X68108; CAA48228.1; -
DR EMBL; Z48502; CAA88413.1; -
DR PIR; S28537; S28537
DR PIR; S28925; S28925
DR PIR; A44402; A44402
DR SGD; S0004650; NUP116.
DR KW Nuclear protein; Transport; Repeat.
FT DOMAIN 205 715 37 X 6 AA APPROXIMATE REPEATS OF
FT G-L-F-G.
FT CONFLICT 26 26 G -> A (IN REF. 1).
FT CONFLICT 536 536 S -> G (IN REF. 1).
FT CONFLICT 720 720 S -> P (IN REF. 1).
FT CONFLICT 1018 1018 S -> Y (IN REF. 1).
FT CONFLICT 1023 1023 I -> Y (IN REF. 1).
SQ SEQUENCE 1113 AA; 116234 MW; FBAB0B9AEA958213 CRC64;

Query Match 6.7%; Score 155; DB 1; Length 1113;
Best Local Similarity 23.0%; Pred. No. 0.39;
Matches 115; Conservative 60; Mismatches 180; Indels 144; Gaps 26;

Qy 11 SSSPGLF---QSGDNGL-GCHNANSALGQPIDRQIEQAQLLAELLKSLSPQSGN 65
Db 403 SNAGLFGONNQSQNSGLFGQNSSNAFQGP-----QQQGLFGSKPAGGLFGQQQ 455
Qy 66 AATGAGNDQTTGV-----GNAGLNGRK-----GTAGTTQSDSONMLSEMG 108
Db 456 ASTFASGNAQNNSIFGQNQQOOSTGGGLFGQONNQSOPGGLFGQTQNNNQ-----PFG 511
Qy 109 NGLDQAITPDGQGGQIGDNPLLKMLKLIARMMDGSDQDQPGPTGNN----- 158
Db 512 QNGLOQ----PQNNNSLFGAKPTGFGNTSLFSNTTQNSNGI-----SGNNLQQQSGGLFQ 563
Qy 159 ----SASSGTSSSGSPNDLSGKAPSGNSPGNY-SPVSTF-----SPST-----PT 203
Db 564 NKQPPASGGL--FGSKPNTVGGGL--FENNOVANQNNPASTSGGLFGSKPATGSLFGGT 619
Qy 204 SPTSPLDFFSPPTKAAG-----STPVTDHPDPVGSAGIGAGNSVAETSAQA----- 250
Db 620 NST-----APNASSGGIFGSGNNASNTAATTSTGLFKNPVGAG---ASTSAGGLFGN 669
Qy 251 -NOTVLHD-----TITVKAQVFDGKGQTTAGSEL-----GDGGQSENQKP 291
Db 670 NNNSSLNNSNGSTGLFGSNNTSQSTNAGLFQNTSTNTSGGLFSPQSQAQSQN--- 726
Qy 292 LFLEDGASLKNVTMGDDGADGIH-LYGDAKIDNLHVTNVGDEAITVKPNSAGKKS---- 346
Db 727 --ALQQQQQQQRLQIQNNNPYGTNELFSKAVTNTVSPATKIKADERKKASLTNA 784
Qy 347 -----HVEITNSFEHASDKIQLNADTNLSVDNVKAKDFTGTVRTNGQQQGNW 395
Db 785 VKMIPKTLFTAKLKTNNVMDKA-----QIKVDPKLSISIDKKN---QIAISNQEENL 836
Qy 396 D---LNLSHISAECKFSF 411
Db 837 DESILKASELLFNPDKRSF 855
```